Feedback Effects in Stochastic Control Problems with Liquidity Frictions

DISSERTATION

zur Erlangung des akademischen Grades

doctor rerum naturalium
(Dr. rer. nat.)
im Fach Mathematik

eingereicht an der
Mathematisch-Naturwissenschaftlichen Fakultät
der Humboldt-Universität zu Berlin

von
M.Sc. Todor Bilarev

Präsidentin der Humboldt-Universität zu Berlin:
Prof. Dr.-Ing. Dr. Sabine Kunst

Dekan der Mathematisch-Naturwissenschaftlichen Fakultät:
Prof. Dr. Elmar Kulke

Gutachter:
1. Prof. Dr. Dirk Becherer
2. Prof. Dr. Peter Bank
3. Prof. Dr. Bruno Bouchard

eingereicht am: 05.06.2018
Tag der Verteidigung: 11.10.2018

Abstract

In this thesis we study mathematical models of financial markets with a large trader (price impact models) whose actions have transient impact on the risky asset prices. Typically in such models the price process of the risky asset is specified as a function of exogenously given risk factors, e.g. a fundamental price process, and processes that capture the illiquidity effects and are driven by the control action of the large trader. Thus, the prices of the risky asset and the proceeds from trading depend in a complex nonlinear way on his actions, hence leading to modeling and optimization problems with feedback effects.

At first, we study the question of how to define the large trader's proceeds from trading. Starting with absolutely continuous controls we identify the proceeds as a nonlinear integral where the integrator and the integrand both depend on the control. To extend the proceeds functional to general controls, in particular controls with jumps or even of infinite variation, we ask for stability in the following sense: nearby trading activities should lead to nearby proceeds. Our main contribution in this part is to identify a suitable topology on the space of controls, namely the Skorokhod M_1 topology, and to obtain the continuous extension of the proceeds functional from absolutely continuous to general càdlàg (right continuous with left limits) controls. Apart from identifying the asymptotically realizable proceeds, we demonstrate by examples how continuity properties are useful to solve different stochastic control problems on optimal liquidation.

Secondly, we solve the optimal liquidation problem in a multiplicative price impact model where liquidity is stochastic in that the volume effect process, which determines the inter-temporal resilience of the market, is taken to be stochastic, being driven by own random noise. The optimal control is obtained as the reflection local time of a diffusion process reflected at a non-constant free boundary. To solve the HJB variational inequality and prove optimality, we need a combination of probabilistic arguments and calculus of variations methods, involving Laplace transforms of inverse local times for diffusions reflected at elastic boundaries.

In the second half of the thesis we study the hedging problem for a large trader. We solve the problem of superhedging for European contingent claims in a multiplicative impact model using techniques from the theory of stochastic target problems. The minimal superhedging price is identified as the unique viscosity solution of a semi-linear pde (possibly with constraints on the gradient), whose nonlinearity is governed by the transient nature of price impact. When a sufficiently regular classical solution exists, a replicating strategy is described by the respective partial derivatives.

Finally, we extend our consideration to multi-asset models where cross-impact is an important new aspect. Requiring stability leads to strong structural conditions that arbitrage-free models with cross-impact should satisfy. These conditions turn out to be crucial for identifying the proceeds functional for a general class of strategies. As an application, the problem of superhedging with cross-impact in additive price impact models is solved.

Zusammenfassung

In dieser Arbeit untersuchen wir mathematische Modelle für Finanzmärkte mit einem großen Händler, dessen Handelsaktivitäten transienten Einfluss auf die Preise der Anlagen haben. Typisch für solche Modelle ist, dass der Preisprozess festgelegt ist als eine Funktion von exogenen Faktoren, z.B. ein fundamentaler Preisprozess, sowie von Prozessen, welche die Illiquiditätseffekte abbilden und deren Dynamik von der Strategie des großen Händlers getrieben wird. Somit hängen die Preise und seine Erlöse in einer komplexen nichtlinearen Weise von seinem Handeln ab, was zu mathematisch herausfordernden Modellierungs- und Optimierungsproblemen mit Feedback-Effekten führt.

Zuerst beschäftigen wir uns mit der Frage, wie die Handelserlöse des großen Händlers definiert werden sollen. Wir identifizieren die Erlöse zunächst für absolutstetige Strategien als nichtlineares Integral, in welchem sowohl der Integrand als der Integrator von der Strategie abhängen. Um die Definition des Funktionals für die Handelserlöse auf allgemeinere Strategien zu erweitern, insbesondere auf Strategien mit Sprüngen und von unendlicher Variation, argumentieren wir mit einem Stabilitätsanzatz wie folgt: ähnliche Handelsaktivitäten sollten ähnliche Erlöse liefern. Unserere Hauptbeiträge sind hier die Identifizierung der Skorokhod M_1 Topologie als geeigneter Topologue auf dem Raum aller Strategien sowie die stetige Erweiterung der Definition für die Erlöse von absolutstetigen auf càdlàg (rechtsstetig mit linken Limiten) Kontrollstrategien. Verschiede Beispiele stochastischer Kontrollprobleme zeigen, wie die Stetigkeitseigenschaften von Nutzen sind.

Weiter lösen wir ein Liquidierungsproblem in einem multiplikativen Modell mit Preiseinfluss, in dem die Liquidität stochastisch ist in dem Sinne, dass der Volumen-Effekt-Prozess, der die intertemporale Anpassungsfähigkeit des Marktes bestimmt, eine stochastische Dynamik hat. Die optimale Strategie wird beschrieben durch die Lokalzeit für Reflektion einer Diffusion an einer nicht-konstanten Grenze. Um die HJB-Variationsungleichung zu lösen und Optimalität zu beweisen, wenden wir probabilistische Argumente und Methoden aus der Variationsrechnung an, darunter Laplace-Transformierte von Lokalzeiten für Reflektion an elastischen Grenzen.

In der zweiten Hälfte der Arbeit untersuchen wir die Absicherung (Hedging) für Optionen. Wir lösen das Superhedging-Problem für Europäische Optionen in einem multiplikativen Preis-Impakt-Modell mit Techniken aus der Theorie für stochastische Zielprobleme. Der minimale Superhedging-Preis ist die Viskositätslösung einer semi-linearen partiellen Differenzialgleichung (gegebenfalls mit Gradientenbeschränkungen), deren Nichtlinearität von dem transienten Preiseinfluss abhängt. Falls eine klassische Lösung der Gleichung mit genügender Glattheit existiert, wird durch sie eine replizierte Hedging-Strategie beschrieben.

Schließlich erweitern wir unsere Analyse auf Hedging-Probleme in Märkten mit mehreren riskanten Anlagen, wobei wechselseitiger Preis-Impakt wichtig wird. Stabilitätsargumente führen zu strukturellen Bedingungen, welche für ein arbitragefreies Modell mit wechselseitigem Preis-Impakt gelten müssen. Zudem ermöglichen es jene Bedingungen, die Erlöse für allgemeine Strategien unendlicher Variation in stetiger Weise zu definieren. Als Anwendung lösen wir das Superhedging-Problem in einem additiven Preis-Impakt-Modell mit mehreren Anlagen.

Contents

1 Introduction

The goal of this chapter is to explain the key concepts and problems in the thesis, to embed our approach in the literature and to introduce and outline the main results from the remaining chapters.

Liquidity and Price impact

One crucial assumption in the classical models from financial mathematics is that trading actions of market participants do not have effect on the dynamics of the asset prices. Recently there has been a lot of work on relaxing this assumption by also considering *liquidity frictions*. Liquidity in financial markets refers to either the ease with which assets can be bought and sold, or the ability to trade without triggering important changes in the evolution of asset prices. However, in many cases due to limited supply and demand trading large volumes moves prices, typically in unfavorable direction. In such cases liquidity becomes an important friction that has to be considered when making trading decisions.

A common approach to model illiquidity (lack of liquidity), typical for the so-called large trader models (sometimes also referred to as price or market impact models), is by exogenously specifying the *price impact* from trading actions. The majority of literature on price impact can be divided into two streams. In the first, the impact from trading has two components: temporary (or sometimes referred to as instantaneous), that only affects the current trade and does not trigger changes in the future evolution of prices, and thus can be also seen as (non-proportional) transaction costs, and (purely) permanent, that affects the future evolution of the risky asset price in a persistent way. First models of this type were proposed in [BL98, AC01]. In continuous time, temporary impact in such models is typically measured in terms of the rate of trading and hence only absolutely continuous trading is allowed. The models in [BB04] and [ÇJP04] fall also into this category but a larger class of trading programs like semimartingales are feasible: [BB04] have only permanent impact while [ÇJP04] consider only temporary impact where (instantaneous) costs are specified for block trades through a supply curve.

The second stream of literature takes a step further and incorporates the well-observed empirical fact that a substantial part of the permanent impact may decay in time, i.e. impact can be also *transient*. One of the first models to incorporate transient impact is that of [OW13], later extended in [AFS10, PSS11]. In these works, the price impact is derived from the presence of a limit order book (LOB) and is a result of demand/supply imbalances triggered by trades. These imbalances recover, for example by new orders arriving in the market, thus rendering the price impact as transient. In mathematical terms, the volume imbalance is described by a *market impact process* $Y = Y^\Theta$ that is driven by the trading strategy Θ of the large trader, and evolves according to a mean-reverting differential equation like

$$\mathrm{d}Y_t = -h(Y_t)\,\mathrm{d}t + \mathrm{d}\Theta_t,$$

where Θ_t is the amount of risky assets held by the trader at time t and h is the resilience function that models the impact relaxation effects. The price of the risky asset in the aforementioned works is then basically of an additive form like

$$S_t = \overline{S}_t + f(Y_t^\Theta),$$

where $\overline{S}$ is the so called *unaffected* (or fundamental) price process that would prevail in absence of market impact, and f is a suitable impact function that is also related to the shape of a LOB. Subject to suitable properties of the functions f and h, asset sales (buys) are depressing (increasing) in a transient way the level of market impact Y_t and thus also the actual price S_t, with some finite rate of resilience.

Overview of the Thesis

In the first part of this thesis, we build upon the transient price impact models in the last category and postulate that the price of the risky asset is of the general form

$$S_t = g(\overline{S}_t, Y_t^\Theta),$$

where g is a suitable function, $\overline{S}$ and Y^Θ are the fundamental price and impact processes respectively. The first problem that arises is how to specify the *proceeds*, being the negative costs for financing the selling and buying in the risky asset. This object is fundamental for the optimization problems one faces in the context of price impact. We take the following point of view: *"nearby" tradings should yield "nearby" proceeds.* Thereby, the proceeds functional should be continuous on the space of strategies with respect to a topology that characterizes "nearby". For instance, a block trade should yield approximately the same proceeds as if we were to split this block trade into smaller block trades and execute them quickly one after the other. As it turns out, the Skorokhod M_1 topology on the space of càdlàg paths captures exactly such reallocations in space and time. Our analysis starts from the observation that under minimal assumptions on how one defines the proceeds from a block trade, the proceeds from absolutely continuous strategies seen as the (uniform) limit of simple strategies, i.e. a finite number of block trades, are

$$L_t(\Theta) = -\int_0^t g(\overline{S}_u, Y_u^\Theta)\,\mathrm{d}\Theta_u.$$

Moreover, the set of absolutely continuous functions form a dense set in the space of càdlàg functions with respect to the Skorokhod M_1 topology. Thus, the question of defining the proceeds for a large class of strategies boils down to identifying the continuous extension (in the M_1 topology) of the functional L from the set of absolutely continuous to càdlàg controls. This we do in Chapter 2. In particular, the proceeds from block trades and semimartingale trading strategies are identified as a continuous extension from those from continuous finite-variation strategies.

Our key example is that of *multiplicative impact* where $g(\overline{S}, Y) = \overline{S}f(Y)$ for a suitable positive function f. From conceptual point of view, such specification guarantees positivity of asset prices as long as the fundamental price $\overline{S}$ is positive, overcoming a theoretical drawback of the aforementioned models with additive structure. For multiplicative impact models we show absence of arbitrage opportunities and demonstrate how continuity of the proceeds functional is useful to prove existence result on optimal strategies in the optimal liquidation problem, that is the problem of how to optimally execute a large trade. Moreover, we demonstrate the scope of our analysis by considering different extensions of our setup, for example when impact could also be partially instantaneous or partially persistent, where we are also able to identify the proceeds for general trading strategies.

In most of the literature on the optimal execution problem under price impact the optimal liquidation strategies are typically *static*, that is the trade schedule can be determined before the trading begins and is not modified by the new information revealed. In contrast, liquidity becomes a risk when some of its aspects are stochastic, and one expects that optimal trading behavior should be *adaptive* to its random changes. To study how optimal liquidation strategies behave under stochastic liquidity, we extend in Chapter 3 our multiplicative impact model from Chapter 2 by considering volume imbalances that have their own stochasticity. In this setup, we solve the optimal liquidation problem over the infinite time horizon. The optimal trading strategy turns out to be of local-time type and is described by reflecting the market impact process (modeling the stochastic volume imbalance) at a non-constant free boundary. The latter is described explicitly up to the solution of an ode parameterized by the size of the remaining position. In particular, the form of the boundary reveals that the more assets are left to be sold, under less favorable market conditions the large trader should sell. Apart from its application context, the analysis of the resulting optimization problem is interesting because it combines calculus of variations techniques with new probabilistic results, involving diffusions reflected at non-constant elastic boundaries and the Laplace transforms of their inverse local times.

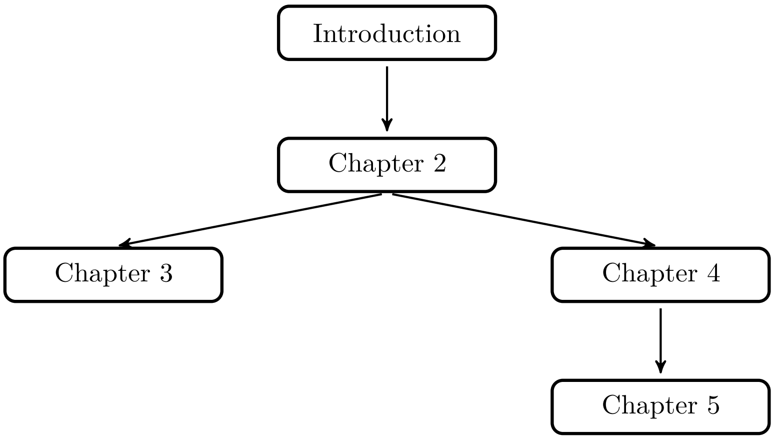
Another fundamental problem in Mathematical Finance studied also in the context of price impact is that of pricing and hedging derivatives. For the typical approaches and an extensive overview of the literature before the year 2011 we refer the reader to [GRS11]. More recently, there have been a lot of work on hedging with (purely) temporary impact that typically leads to linear quadratic optimal tracking problems. For more details and literature overview in this direction, see [Voß17]. The subject of interest in the second part of this thesis, namely Chapters 4 and 5, is the problem of pricing and hedging of European contingent claims under transient and permanent impact.

In the context of price impact the hedging problem becomes more complex since the large trader by hedging the option influences the price of the underlying, which on the other hand defines the option's payout at maturity. As a consequence, this may give the large trader an incentive to influence in his favor the price, and thus the payout. We restrict the possibility of manipulations by distinguishing between physical and cash delivery part in the option's payoff and requiring that the physical part shall be delivered exactly. Thus, doing trades shortly before maturity that shall be unwound right after delivery, hence influencing the option's payout to favor the large trader, will not be allowed. In particular, a hedging strategy for a payoff with pure cash delivery part should be a round trip, i.e. it begins and ends with zero shares in the underlying, while for a payoff with non-trivial physical delivery part should be such that the net traded assets at maturity will be exactly enough to meet the physical delivery part. Thus, hedging strategies for European contingent claims with physical delivery part will be different from these with pure cash delivery part, and as it will turn out their respective prices will also differ.

In Chapter 4, we solve the pricing and hedging problem in a single-asset impact model with transient multiplicative impact, by following the *super-replication* approach. Using techniques for stochastic target problems, we characterize the *minimal superhedging price* as the unique viscosity solution of a semi-linear partial differential equation (pde for short). Here, the non-linearity in the pde is governed by the resilience and price impact functions. In a case study of exponential impact function we show how constraints on the *"delta"* (first-order spacial derivative) are induced in order to have a well-posed pde. If the pricing pde admits a sufficiently regular classical solution, then an optimal hedging strategy can be constructed and has the following structure: there is one initial and one terminal block trade, and continuous trading in between, being typically of infinite variation. We also demonstrate how the pricing and hedging problem drastically changes in the following seemingly mild modification: the price impact from initial and terminal trades is omitted. This specification corresponds to the so-called covered options and the pricing pde turns out to be of different nature: it is fully non-linear with a singularity on the second-order spacial derivative, hence inducing *"gamma"* constraints.

In the literature price impact is mostly investigated in a single-asset setup, like we did so far. In the last part of this thesis, Chapter 5, we extend our analysis to multi-asset models. With multiple assets, a new form of impact that can become relevant which is the effect of a trade in one asset on the price dynamics of another asset, that is called *cross-impact.* We consider a general price specification with both permanent and transient impact. Assuming absence of instantaneous round trips that give positive proceeds, we derive structural conditions on the price impact function, namely that it is a gradient field. In turn, these structural conditions suffice for the ideas from Chapter 2 to carry over here and thus to identify the proceeds for general càdlàg trading strategies as the continuous extension from simple strategies. As a consequence, we identify the wealth process from general self-financing tradings. Thereby, we are in a position to study the problem of pricing and hedging of contingent claims. In an additive price impact specification with cross-impact, we characterize the minimal superhedging price as the unique viscosity solution of a semi-linear pde. The non-linearity is governed by the transient nature of impact, similarly to our findings from Chapter 4. In particular, the persistent permanent impact is irrelevant for the pricing pde, however the hedging strategy is affected by it.

In the view of our contributions above, the main theme of this thesis is transient price impact models and optimization problems with feedback effects. The applications are the optimal liquidation and the pricing and hedging of contingent claims in illiquid markets. Each chapter is written in a self-contained way and results from the previous chapters are stated precisely. The interdependence of the chapters is as follows:



In what follows we introduce in detail the results of the next chapters and explain how they complement the existing literature.

Stability for gains from large investors' strategies in $\mathrm{M}_1/\mathrm{J}_1$ topologies (Chapter 2)

Defining proceeds for general strategies by continuous extension

A classical theme in the theory of stochastic differential equations is how stably the solution process behaves, as a functional of its integrand and integrator processes, see e.g. [KP96] and [Pro05, Chapter V.4]. A typical question is how to extend such a functional sensibly to a larger class of input processes. Continuity is a key property to address such problems, cf. e.g. the canonical extension of Stratonovich SDEs by Marcus [Mar81].

In singular control problems for instance, the non-linear objective functional may initially be only defined for finite variation or even absolutely continuous control strategies. Existence of an optimizer might require a continuous extension of the functional to a more general class of controls, e.g. semimartingale controls for the problem of hedging. Herein the question of which topology to embrace arises, and this depends on the problem at hand, see e.g. [Kar13] for an example of utility maximization in a frictionless financial market where the Emery topology turns out to be useful for the existence of an optimal wealth process. For our application we need suitable topologies on the Skorokhod space of càdlàg functions. The two most common choices here are the uniform topology and Skorokhod J_1 topology; they share the property that a jump in a limiting process can only be approximated by jumps of comparable size at the same time or, respectively, at nearby times. But this can be overly restrictive for such applications, as we have in mind, where a jump may be approximated sensibly by many small jumps in fast succession or by continuous processes such as Wong-Zakai-type approximations. The M_1 topology by Skorokhod [Sko56] captures such approximations of *unmatched jumps*. We will take this as a starting point to identify the relevant non-linear objective functional for càdlàg controls as a continuous extension from (absolutely) continuous controls. See [Whi02] for a profound survey on the M_1 topology.

We demonstrate how the old subject of stability of SDEs with jumps, when considered with respect to the M_1 topology, has applications for recent problems in mathematical finance. Our application context is that of an illiquid financial market for trading a single risky asset. A large investor's trading causes transient price impact on some exogenously given fundamental price which would prevail in a frictionless market. Such could be seen as a non-linear (non-proportional) transaction cost with intertemporal impact also on subsequent prices. Our framework is rather general. It can accommodate for instance for models where price impact is basically additive, see Example 2.1.1; Yet, some extra care is required here to ensure M_1 continuity, which can actually fail to hold in common additive models that lack a monotonicity property and positivity of prices, cf. Remark 2.2.9. An original aspect of our framework is that it also permits for multiplicative impact which appears to fit better to multiplicative price evolutions as e.g. in models of Black-Scholes type, cf. [BBF17a, Example 5.4]; In comparison, it moreover ensures positivity of asset prices, which is desirable from a theoretical point of view, relevant for applications whose time horizon is not short (as they can occur e.g. for large institutional trades [CL95, KMS17], or for hedging problems with longer maturities like in Chapter 4).

The large trader's feedback effect on prices causes the proceeds (negative expenses) to be a non-linear functional of her control strategy for dynamic trading in risky assets. Having specified the evolution for an affected price process at which trading of infinitesimal quantities would occur, one still has, even for a simple block trade, to define the variations in the bank account by which the trades in risky assets are financed, i.e. the so-called self-financing condition. Choosing a seemingly sensible, but ad-hoc, definition could lead to surprising and undesirable consequences, in that the large investor can evade her liquidity costs entirely by using continuous finite variation strategies to approximate her target control strategy, cf. Example 2.2.2. Optimal trade execution proceeds or superreplication prices may be only approximately attainable in such models. Indeed, the analysis in [BB04, ÇJP04] shows that approximations by continuous strategies of finite variation play a particular role. This is, of course, a familar theme in stochastic analysis, at least since Wong and Zakai [WZ65]. However, in the models in [BB04, ÇJP04] the aforementioned strategies have zero liquidity costs, permitting the large trader to avoid those costs entirely by simply approximating more general strategies. This appears not desirable from an application point of view, and it seems also mathematically inconvenient to distinguish between proceeds and asymptotically realizable proceeds. To settle this issue, a stability analysis for proceeds for a class of price impact models should address in particular the M_1 topology, in which continuous finite variation strategies are dense in the space of càdlàg strategies (in contrast to the uniform or J_1 topologies), see Remark 2.2.5.

We contribute a systematic study on stability of the proceeds functional. Starting with an unambiguous definition (2.4) for continuous finite-variation strategies, we identify the approximately realizable gains for a large set of controls. A mathematical challenge for stability of the stochastic integral functional is that both the integrand and the integrator depend on the control strategy. The main Theorem 2.2.7 in this chapter shows continuity of this non-linear controlled functional in the uniform, J_1 and M_1 topologies, in probability, on the space of (predictable) semimartingale or càdlàg strategies which are bounded in probability. A particular consequence is a Wong-Zakai-type approximation result, that could alternatively be shown by adapting results from [KPP95] on the Marcus canonical equation to our setup, cf. Section 2.2.3. Another direct implication of M_1 continuity is that proceeds of general (optimal) strategies can be approximated by those of simple strategies with only small jumps. Whereas the former property is typical for common stochastic integrals, it is far from obvious for our non-linear controlled SDE functional (2.15).

The topic of stability for the stochastic process of proceeds from dynamically trading risky assets in illiquid markets, where the dynamics of the wealth and of the proceeds for a large trader are non-linear in his strategies because of his market impact, is showing up at several places in the literature. But the mathematical topic appears to have been touched mostly in-passing so far. The focus of few notable investigations has been on the application context and on different topologies, see e.g. [RS13, Proposition 6.2] for uniform convergence in probability (ucp). In [LS13, Lemma 2.5] a cost functional is extended from simple strategies to semimartingales via convergence in ucp. [Roc11, Definition 2.1] and [ÇJP04, Section A.2] use particular choices of approximating sequences to extend their definition of self-financing trading strategies from simple processes to semimartingales by limits in ucp. Trading gains of semimartingale strategies are defined in [BLZ16, Prop. 1.1–1.2] as L^2-limits of gains from simple trading strategies via rebalancing at discrete times and large order split. In contrast, we contribute a study of M_1-, J_1- and ucp-stability for general approximations of càdlàg strategies in a class of price impact models with transient impact (2.3), driven by quasi-left continuous martingales (2.1).

As a further contribution, and also to demonstrate the relevance and scope of the theoretical results, we discuss in the case of multiplicative impact a variety of examples where continuity properties play a role. In Section 2.4.2 we establish existence of an optimal monotone liquidation strategy in finite time horizon using relative compactness and continuity of the proceeds functional in M_1. Section 2.4.3 shows how to solve the optimal liquidation problem in infinite time horizon with non-negative bounded semimartingale strategies by approximating their proceeds via bounded variation strategies, here the M_1-stability being needed. Section 2.4.4 incorporates partially instantaneous recovery of price impact to our model, while in Section 2.4.5 we consider permanent impact as well. Herein, the M_1 topology plays the key role to identify (asymptotically realizable) proceeds as a continuous functional. Last but not least, Section 2.3 proves absence of arbitrage for the large trader within a fairly large class of trading strategies.

Optimal Liquidation under Stochastic Liquidity (Chapter 3)

A singular stochastic control problem

A typical stochastic optimal control problem in models of illiquid markets is a large trader (the controller) who optimizes her trading strategy such as to balance some trading objective against her adverse price impact, which causes (non-proportional) cost from illiquidity. In the majority of literature on price impact models the inter-temporal impact is typically a deterministic function of the strategy of the (single) large trader. In reality, we would rather expect some aspects of market liquidity (where [Kyl85] has distinguished resilience, depth and tightness) to vary stochastically over time, and a sophisticated trader to adapt her optimal strategy accordingly. Even for the extensively studied problem of optimal liquidation, there are relatively few recent articles on models in continuous time where the optimal liquidation strategy is adaptive to random changes in liquidity, cf. [Alm12, LS13, FSU17, GHS18, GH17].

We consider a model where temporary market imbalances involve *own stochasticity*. Price impact is transient, i.e. it could be persistent but eventually vanishes over time. Moreover, it is non-linear, corresponds to a general shape for the density of the limit order book (see Remark 3.1.3), and is multiplicative to ensure positive risky asset prices. More precisely, our price process $S = (S_t)_{t\geq 0} = (f(Y_t)\overline{S}_t)_{t\geq 0}$ observed in the market deviates by a factor $f(Y_t)$ from the fundamental price $\overline{S}_t$ that would prevail in the absence of large traders. The impact function f is positive and increasing and thus the multiplicative structure ensures that prices stay positive, in contrast to the additive models where a conceptual drawback is that negative asset prices can occur with (small) positive probability. Our stochastic impact process Y is of a controlled Ornstein-Uhlenbeck (OU) type, namely it is driven by a Brownian motion and the large trader's holdings in the risky asset (see eq. (3.3) below). The mean-reversion of Y models the transience of impact. Analogously to [AFS10, PSS11], the impact function f can be linked to the shape of a limit order book (LOB) and Y may be understood as a volume effect process describing the (temporal) imbalance in the LOB, see Remark 3.1.3. The additional noise in Y gives a stochastic LOB, or it can be seen as the accumulated effect from other non-strategic large traders, see Remark 3.1.4.

For our multiplicative model with transient impact, we take the fundamental price $\overline{S}$ to be an exponential Brownian motion and permit for non-zero correlation with the stochastic volume effect process Y. In this setup, we study the optimal liquidation problem for infinite time horizon as a singular stochastic control problem of finite fuel type and construct its explicit solution. Our main result in this chapter, Theorem 3.2.1, gives the optimal strategy as the local time process of a diffusion reflected obliquely at a curved free boundary in $\mathbb{R}^2$, the state space being the impact level and the holdings in the risky asset. In contrast to notable research on adaptive strategies (in different models) by [SS09, LA11], the stochasticity of our strategy arises from its adaptivity to the transient component of the price dynamics and is of local time type. Moreover, different from models with additive price impact where the martingale part of the fundamental price is irrelevant for a risk-neutral trader, here the volatility of $\overline{S}$ is important in that it is a parameter in the equation for the free boundary, cf. Remark 3.2.3.

We solve the singular control problem by explicitly constructing the value function as a classical solution of the Hamilton-Jacobi-Bellmann variational inequality. Our verification arguments differ from a more common approach (outlined in Remark 3.5.4) since we were not able to verify the optimality more directly, due to the technical complications arising from the implicit nature of the eigenfunctions of the infinitesimal generator for the Ornstein-Uhlenbeck process (see Remark 3.5.8). In contrast, we first restrict the set of optimization strategies to those described by diffusions reflected at monotone boundaries, and optimize over the set of possible boundaries. To be able to apply methods from calculus of variations, we derive an explicit formula (eq. (3.17)) for the Laplace transform of the inverse local times of diffusions reflected at elastic boundaries, i.e. boundaries which vary with the local time that the reflected process has spent at the boundary, and employ a change of coordinates. By solving the calculus of variations problem, we construct the candidate optimal free boundary and, moreover, show (one-sided) local optimality in the sense of Theorem 3.4.6. The latter is crucial for our verification of optimality.

Superhedging with transient impact of non-covered and covered options (Chapter 4)

A stochastic target problem

The derivation of the Black-Scholes formula for the price of European options is fundamental for the development of the no-arbitrage theory for pricing and hedging in Mathematical Finance. Relaxing its main assumptions, for instance that the market is frictionless, has been subject to a lot of research since then. In this chapter, we study the *superhedging problem* in a market model where the actions of a trader have impact on the price of the risky asset. This is in contrary to the assumption of perfect liquidity in the Black-Scholes model, where it is assumed to be possible to buy or sell arbitrary quantities of the risky asset without affecting its dynamics. The impact mechanism from the trader's actions that we consider leads to non-trivial intertemporal effects on the drift and volatility of the price process, which lessen in time, meaning that price impact is transient, and thus the price would recover towards the reference price from the Black-Scholes model. The superhedging problem in this setup features non-trivial feedback effects in the following sense, cf. [SW00]: A hedging strategy directly influences the price of the underlying risky asset with respect to which the option's payoff at maturity is defined.

While pricing and hedging of derivatives in models with price impact has been already studied before, cf. [FP11, BSV16] and the references therein, the prevailing literature considers models where impact is instantaneous, meaning that it will not affect the future price but only current trading and hence can be seen as non-proportional transaction costs, or purely permanent, that is when the (full) future evolution of the price process is affected by current trade and there are no relaxation effects. In contrast, the focus here is on the *transient* nature of impact and how, if at all, it affects the minimal superhedging price and the hedging strategy for a European option. More precisely, we consider the multiplicative impact model from Chapter 2 where the large trader's actions, modeled by the evolution of his risky asset position Θ, affect the market impact Y^Θ measuring the volume imbalance in a static LOB. The process Y^Θ is mean-reverting, meaning that these imbalances tend to cease in time, and thus the impact is transient. The relative deviation from the fundamental (or unaffected) price $\overline{S}$, that should prevail in the absence of price impact, is the positive (multiplicative) factor $f(Y^\Theta)$ for an impact function f.

In contrast to the classical liquid models, because of price impact the transfer of funds between the riskless and the risky asset accounts does not come for free: apart from possibly paying liquidity costs, selling or buying the risky asset directly affects its price. That is why the way the option's payoff is specified at maturity is highly relevant for the problem. In fact, as already observed recently in [BLZ16] and [BLZ17] for a related impact model with permanent impact, considering or disregarding the impact from initial and terminal trades (*non-covered* or *covered options* respectively) leads to completely different problems with different in nature pricing equations (quasi-linear versus fully non-linear for their setup). In the case of non-covered options, it is also important to distinguish between physical delivery and cash delivery, that is whether or not a position of the risky asset should be delivered at maturity. Indeed, it has been emphasized in [GP17] that physical or cash delivery lead to different (in their case utility indifference) prices, in a model with temporary and permanent impact.

In this chapter we address both problems of pricing and hedging of covered and non-covered options and show that resilience plays a non-negligible role for the pricing equation in the non-covered option case. In this case, we reformulate the pricing problem as a stochastic target problem and derive a Dynamic Programming Principle (DPP) along suitably chosen coordinates, which are *the effective price and impact processes*, being the price and impact that would prevail if the trader was to immediately clear his position in the risky asset. Along these coordinates, DPP gives a way to compare (at stopping times) *the instantaneous liquidation wealth* and the minimal superhedging price. This enables us to derive a non-linear pricing equation for the (minimal) superhedging price, under general assumptions on the (non-parametric) impact and resilience functions. The pricing pde turns out to be a semi-linear Black-Scholes pde whose nonlinearity involves the resilience and the impact functions h and f. Moreover, if it has a sufficiently regular solution, an optimal replicating strategy can be constructed. In this case, the hedging strategy will incorporate the transient nature of impact in that it will depend on the level of (effective) impact.

Our analysis is inspired by [BLZ16] where in a model with purely permanent (additive) impact the pricing and hedging of non-covered options is solved as a stochastic target problem. For the present setup, we need to consider an extended state space which in addition includes the level of impact. In particular, our results show that the current deviation of the asset price from the unaffected price is an important new state variable on which the price of the option and the hedging strategies depend non-trivially. In addition, having cash or physical delivery at maturity leads to different boundary conditions for the pricing pde and hence typically different prices. In particular, the superhedging price of a European call option with cash delivery is smaller than that with physical delivery.

For the results so far we consider general impact functions f under the assumption that f is bounded away from zero and infinity, meaning that the large trader cannot indefinitely affect the price of the risky asset as much as he wants but only by at most a fraction. Departing from this assumption, we consider in Section 4.4.2 the case where f is the exponential function. There we see how our analysis could also be applied to derive a pricing pde for the well-posedness of which we introduce *Delta constraints* on the admissible hedging strategies. In this setup, it turns out that the pricing pde for a typical European option, that means whose payoff is given by a function of the price of the underlying only, reduces to the Black-Scholes pde with gradient constraints.

That the form of the resilience is immaterial for the price but not for the hedging strategy of covered options was already pointed out in [BLZ17, Section 3] (in a different setup with additive impact though). We sketch in Section 4.7 how a similar analysis carries over to our setup and derive a singular pricing pde the analysis of which induces *Gamma constraints*. In contrast to the results in [BLZ17], it turns out that the current deviation of the asset price from the unaffected price becomes a relevant new state variable for describing the solution.

Cross-impact and hedging in multi-asset price impact models (Chapter 5)

Modeling and application to pricing and hedging

In this chapter, we extend our single-asset setup to multi-asset impact models. When considering more than one illiquid assets, a new form of impact becomes relevant, namely *cross-impact*. This is the price impact that trades in one asset have on the price dynamics of another asset. Multi-asset impact for optimal trading has been recently investigated only in a few papers and mainly in the context of optimal execution, see the literature overview below. A key finding there is that because of cross-impact, syncronized trading in multiple assets might substantially reduce liquidity costs. It was demonstrated in e.g. [Sch16, TWG17] that even when the goal is to trade a single asset, it might be optimal to do intermediate trading in other assets. Indeed, trading in other assets provide an opportunity for risk reduction through diversification and cross-impact moreover may give additional benefits in reducing execution costs. Thus, utilizing properly these two effects can lead to reduction in execution costs.

We extend the single-asset specification from Chapter 2 to a multi-asset setup with cross-impact. At first, we postulate that the price of the risky asset is a function g of a multivariate fundamental price process $\overline{S}$, the impact process Y and the trading strategy Θ, i.e. we consider both transient and permanent impact. Here $\overline{S}$ captures the exogenous correlated risks between different assets, while the illiquidity (cross-)effects will be captured by a multidimensional process Y mean reverting dynamics, reflecting temporary supply and demand imbalances, extending the one-dimensional impact process so far. In such a setup, we would like to study how these effects jointly influence the trading behavior of a large trader who wants to hedge a contingent claim. To do an analysis like the one in Chapter 4, we need to specify the proceeds from general càdlàg strategies by continuously extending the proceeds functional. Our analysis starts from the observation that a sensible model specification should not allow for profitable asymptotically instantaneous round trips, being limits (as time interval goes to zero) of absolutely continuous round trips executing in a small time interval that yield positive proceeds. As it turns out, this basic assumption implies strong structural properties for the impact function g, requiring a gradient field structure for g. This permits to extend methods from Chapter 2 in order to identify the proceeds for general càdlàg strategies as the continuous extension of the proceed functional for simple strategies. That in turn allows us to extend the analysis from Chapter 4 for the problem of pricing and hedging of contingent claims by the large trader to the multi-asset case. In an additive impact model, we characterize the minimal superhedging price as the viscosity solution of a semi-linear pde, where the non-linearity is governed by the transient component of impact. As a consequence of our analysis in this simple specification, if the option's payoff is a function only of the price of the risky assets (and do not depend on the level of impact or risky assets' holdings), the minimal superhedging price coincides with the friction-less price. A hedging strategy however needs to account for all components of impact.

Related literature on optimal trading in multi-asset price impact models.

Most of the literature on optimal trading with price impact considers one asset. There are a few recent papers on multi-asset models with cross impact that are mostly considered in the context of the optimal trade execution problem. For this problem typically only finite-variation or even absolutely continuous strategies are required. In particular, [SST10] consider a cross-impact model with both permanent and temporary impact and restrict attention to absolutely continuous strategies only. [Sch16] considers a cross-impact model with purely temporary impact (measured in the rate of trading) and permanent cross-impact, and demonstrate how cross-impact could imply that trading in two assets reduces costs even when the objective is to liquidate the position in only one of the assets. Recently in [SL17] no arbitrage implications yield structural conditions on a cross-impact model with both transient and instantaneous impact that can be directly verified on data. It turns out that a necessary (but not always sufficient) condition for absence of so-called price manipulations, that are round trips with negative expected costs, is the symmetry of cross-impact. Based on a recent empirical study, [MBEB17] propose a transient cross-impact model justified and solve the optimal liquidation problem. It is shown that in the presence of cross-impact, synchronized trading is essential for reducing execution costs.

In discrete time, the paper [AKS16] considers both permanent and transient impact and general decay kernels which model the intertemporal relaxation of price impact. Their main contribution is to determine properties of the decay kernel that will lead to well-behaved optimal execution strategies. [TWG17] consider a model similar to our additive impact model being motivated by market microstructure (trading through limit order books) and thus is presented as a multi-asset generalization of the Obizhaeva-Wang model [OW13] in discrete time.

While the aforementioned literature considers the problem of optimal execution, the recent paper [GP16] solves the portfolio choice problem for agents with mean-variance preferences. Their case of "purely persistent costs", corresponding to our additive impact example with purely transient impact and exponential decay, requires strategies of infinite variation. The proceeds for such strategies are obtained as the limit of proceeds from discrete-time tradings.

Acknowledgments

First and foremost, I would like to express my deep gratitude to my supervisor Dirk Becherer. Without his guidance, support and encouragement this thesis wouldn't have been possible. He introduced me to the area of probability theory, guided me through my Master's and Ph.D. studies and always encouraged me to attend numerous conferences and workshops in order to present my research.

Secondly, I would like to thank Peter Bank and Bruno Bouchard for readily accepting to co-examine this thesis. Acting as my mentor, Peter Bank always gave me constructive feedback and suggestions on the projects I was involved in. I am indebted to Bruno Bouchard also for hosting me at Université Paris-Dauphine during a three-months stay in Paris and, despite his tight schedule, for visiting us in Berlin.

Regarding my colleagues, I am particularly thankful to Peter Frentrup for patiently working with me on numerous joint projects. Together we wrote four papers that were possible because of his meticulousness. Moreover, I thank my friends and colleagues Randolf Altmayer, Julio Backhoff, David Beßlich, Jakub Chorowski, Guanxing Fu, Jörg Martin, Victor Fenou Nzengang, Alexandros Saplauras, Klebert Kentia Tonleu, Moritz Voß for all the support and encouragement, and all my friends from the BMS that made my time as a PhD student so enjoyable. I also benefited from discussing with Eduardo Abi Jaber, Sergio Pulido, Zhenjie Ren and Xiaolu Tan.

Financial support by the Deutsche Forschungsgemeinschaft via the Research training group 1845 ("Stochastic Analysis with applications in biology, finance and physics") and the Berlin Mathematical School is gratefully acknowledged.

2 Stability for gains from large investors' strategies in $\mathrm{M}_1/\mathrm{J}_1$ topologies

This chapter, which contains our main contributions in extended form from [BBF17b], lays the foundations for the rest of the thesis. The chapter is organized as follows. Section 2.1 sets the model and defines the proceeds functional for finite variation strategies. In Section 2.2 we extend this definition to a more general set of strategies and prove our main Theorem 2.2.7. Section 2.2.3 draws a link to the Marcus canonical equations, stochastic differential equations of Stratonovich type which are stable for Wong-Zakai approximations. In the remaining Sections 2.3 and 2.4 we concentrate on the case of multiplicative impact. We show absence of arbitrage opportunities for the large investor in Section 2.3 as a basis for a sensible financial model. The examples related to optimal liquidation are investigated in Section 2.4. There we also discuss possible extensions of our setup that incorporate stochastic liquidity, partially instantaneous or permanent impact, and show how our analysis could be applied to these cases as well.

2.1 A model for transient multiplicative price impact

We consider a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$. The filtration $(\mathcal{F}_t)_{t\geq 0}$ is assumed to satisfy the usual conditions of right-continuity and completeness, with $\mathcal{F}_0$ being the trivial σ-field. Paths of semimartingales are taken to be càdlàg. Let also $\mathcal{F}_{0-}$ denote the trivial σ-field. We consider a market with a single risky asset and a riskless asset (bank account) whose price is constant at 1. Without activity of large traders, the unaffected (discounted) price process of the risky asset would evolve according to the stochastic differential equation

$$\mathrm{d}\overline{S}_t = \overline{S}_{t-}(\xi_t \,\mathrm{d}\langle M\rangle_t + \mathrm{d}M_t), \qquad \overline{S}_0 > 0, \tag{2.1}$$

where M is a locally square-integrable martingale that is quasi-left continuous (i.e. for any finite predictable stopping time τ, $\Delta M_\tau := M_\tau - M_{\tau-} = 0$ a.s.) with $\Delta M > -1$ and ξ is a predictable and bounded process. In particular, the predictable quadratic variation process $\langle M\rangle$ is continuous [JS03, Thm. I.4.2], and the unaffected (fundamental) price process $\overline{S} > 0$ can have jumps. We moreover assume that $\langle M\rangle = \int_0^\cdot \alpha_s \,\mathrm{d}s$ with density α being bounded (locally on compact time intervals) and whose paths are (locally) Lipschitz, and that the martingale part of $\overline{S}$ is square integrable on compacts. The assumptions on M are satisfied e.g. for $M = \int \sigma \, \mathrm{d}W$, where W is a Brownian motion and σ is a suitably regular bounded predictable process, or for Lévy processes M with suitable integrability and lower bound on jumps.

To model the impact that trading strategies by a single large trader have on the risky asset price, let us denote by $(\Theta_t)_{t\geq 0}$ her risky asset holdings throughout time and Θ_{0-} be the number of shares she holds initially. The process Θ is the control strategy of the large investor who executes $\mathrm{d}\Theta_t$ market orders at time t (buy orders if Θ is increasing, sell orders if it is decreasing). We will assume throughout that strategies Θ are predictable processes. The large trader is faced with illiquidity costs because her trading has an adverse impact on the prices at which her orders are executed as follows. A *market impact process* Y (called volume effect process in [PSS11]) captures the impact from a predictable strategy Θ with càdlàg paths on the price of the risky asset, and is defined as the càdlàg adapted solution Y to

$$\mathrm{d}Y_t = -h(Y_t)\,\mathrm{d}\langle M\rangle_t + \mathrm{d}\Theta_t \tag{2.2}$$

for some initial condition $Y_{0-} \in \mathbb{R}$. We assume that $h : \mathbb{R} \to \mathbb{R}$ is Lipschitz with $h(0) = 0$ and $h(y)\,\mathrm{sgn}(y) \geq 0$ for all $y \in \mathbb{R}$. The Lipschitz assumption on h guarantees existence and uniqueness of Y in a pathwise sense, see [PTW07, proof of Thm. 4.1] and Proposition 2.5.1 below. The sign assumption on h gives *transience* of the impact which recovers towards 0 (if $h(y) \neq 0$ for $y \neq 0$) when the large trader is inactive. The function h gives the speed of resilience at any level of Y_t and we will refer to it as *the resilience function*. For example, when $h(y) = \beta y$ for some constant $\beta > 0$, the market recovers at exponential rate (as in [OW13, AFS10, Løk14]). Note that we also allow for $h \equiv 0$ in which case the impact is permanent as in [BB04]. Clearly, the process Y depends on Θ, and sometimes we will indicate this dependence as a superscript $Y = Y^{\Theta}$. Some of the results in this chapter could be extended with no additional work when considering additional noise in the market impact process, see the discussion in Section 2.4.6, or for less regular density α if the $-h(Y_t)\mathrm{d}\langle M\rangle_t$ term in (2.2) is replaced by e.g. $-h(Y_t)\mathrm{d}t$.

If the large investor trades according to a continuous strategy Θ, the observed price S at which infinitesimal quantities $\mathrm{d}\Theta$ are traded (see (2.4)) is given via (2.2) by

$$S_t := g(\overline{S}_t, Y_t)\,, \tag{2.3}$$

where the *price impact function* $(x, y) \mapsto g(x, y)$ is $C^{2,1}$ and non-negative with g_{xx} being locally Lipschitz in y, meaning that on every compact interval $I \subset \mathbb{R}$ there exists $K > 0$ such that $|g_{xx}(x, y) - g_{xx}(x, z)| \leq K|y - z|$ for all $x, y, z \in I$. Moreover, we assume $g(x, y)$ to be non-decreasing in both x and y. In particular, selling (buying) by the large trader causes the price S to decrease (increase). This price impact is transient due to (2.2).

Example 2.1.1. [BB04] consider a family of semimartingales $(S^{\theta})_{\theta\in\mathbb{R}}$ being parametrized by the large trader's risky asset position θ. In our setup, this corresponds to general price impact function g and $h \equiv 0$, meaning that impact is permanent. A known example in the literature on transient price impact is the additive case, $S = \overline{S} + f(Y)$, where

[OW13] take $f(y) = \lambda y$ to be linear, motivated from a block-shaped limit order book. For generalizations to non-linear increasing $f : \mathbb{R} \to [0, \infty)$, see [AFS10, PSS11]. Note that we require $0 \leq g \in C^{2,1}$ for Theorem 2.2.7, see Remark 2.2.9. A (somewhat technical) modification of the model by [OW13], that fits with our setup and ensures positive asset prices, could be to take $g(\overline{S}, Y) = \varphi(\overline{S} + f(Y))$ with a non-negative increasing $\varphi \in C^2$ satisfying $\varphi(x) = x$ on $[\varepsilon, \infty)$ and $\varphi(\cdot) = 0$ on $(-\infty, -\varepsilon]$ for some $\varepsilon > 0$. A different example, that naturally ensures positive asset prices and will serve as our prime example for Sections 2.3 and 2.4, is multiplicative impact $S = f(Y)\overline{S}$ for f being strictly positive, non-decreasing, and with $f \in C^1$ (to satisfy the conditions on g). Also here, the function f can be interpreted as resulting from a limit order book, see Section 2.4.1.

While impact and resilience are given by general non-parametric functions, note that these are static. Considering such a model as a low (rather than high) frequency model, we do consider approximations by continuous and finite variation strategies to be relevant. To start, let Θ be a continuous process of finite variation (f.v., being adapted). Then, the cumulative proceeds (negative expenses), denoted by $L(\Theta)$, that are the variations in the bank account to finance buying and selling of the risky asset according to the strategy, can be defined (pathwise) in an unambiguous way. Indeed, proceeds over period $[0, T]$ from a strategy Θ that is continuous should be (justified also by Lemma 2.2.1)

$$L_T(\Theta) := -\int_0^T S_u \,\mathrm{d}\Theta_u = -\int_0^T g(\overline{S}_u, Y_u) \,\mathrm{d}\Theta_u. \tag{2.4}$$

Our main task is to extend by stability arguments the model from continuous to more general trading strategies, in particular such involving block trades and even more general ones with càdlàg paths, assuming transient price impact but no further frictions, like e.g. bid-ask spread (cf. Remark 2.3.4). To this end, we will adopt the following point of view: approximately similar trading behavior should yield similar proceeds. The next section will make precise what we mean by "similar" by considering different topologies on the càdlàg path space. It turns out that the natural extension of the functional L from the space of continuous f.v. paths to the space of càdlàg f.v. paths which makes the functional L continuous in all of the considered topologies is as follows: for discontinuous trading we take the proceeds from a block market buy or sell order of size $|\Delta\Theta_\tau|$, executed immediately at a predictable stopping time $\tau < \infty$, to be given by

$$-\int_0^{\Delta\Theta_\tau} g(\overline{S}_{\tau-}, Y_{\tau-} + x) \,\mathrm{d}x, \tag{2.5}$$

and so the proceeds up to T from a f.v. strategy Θ with continuous part Θ^c are

$$L_T(\Theta) := -\int_0^T g(\overline{S}_u, Y_u) \,\mathrm{d}\Theta^c_u - \sum_{\substack{\Delta\Theta_t \neq 0 \\ 0 \leq t \leq T}} \int_0^{\Delta\Theta_t} g(\overline{S}_{t-}, Y_{t-} + x) \,\mathrm{d}x. \tag{2.6}$$

Note that a block sell order means that $\Delta\Theta_t < 0$, so the average price per share for this trade satisfies $S_t \leq -\frac{1}{\Delta\Theta_t} \int_0^{\Delta\Theta_t} g(\overline{S}_t, Y_{t-} + x) \,\mathrm{d}x \leq S_{t-}$. Similarly, the average price per share for a block buy order, $\Delta\Theta_t > 0$, is between S_{t-} and S_t. The expression in (2.5) could be justified from a limit order book perspective for some cases of g, as noted in Example 2.1.1, see also Section 2.4.1. But we will derive it in the next section using stability considerations.

Remark 2.1.2. The aim to define a model for trading under price impact for general strategies is justified by applications in finance, which encompass trade execution, utility optimization and hedging. While also e.g. [BB04, BR17, ÇJP04] define proceeds for semimartingale strategies, their definitions are not ensuring continuity in the M_1 topology, in contrast to Theorem 2.2.7. Another difference to [BB04, BR17] is that our presentation is not going to rely on non-linear stochastic integration theory due to Kunita or, respectively, Carmona and Nualart.

2.2 Continuity of the proceeds in various topologies

In this section we will discuss questions about continuity of the proceeds process $\Theta \mapsto L_\cdot(\Theta)$ with respect to various topologies: the ucp topology and the Skorokhod J_1 and (in particular) M_1 topologies. Each one captures different stability features, the suitability of which may vary with application context.

Let us observe that for a continuous bounded variation trading strategy Θ the proceeds from trading should be given by (2.4). To this end, let us make just the assumption that

a block order of a size Δ at some (predictable) time t is executed at some average price per share which is between $S_{t-} = g(\overline{S}_t, Y_{t-})$ and $g(\overline{S}_t, Y_{t-} + c\Delta)$ (2.7)

for some constant $c \geq 0$. The assumption looks natural for $c = 1$ where $Y_t = Y_{t-} + c\Delta$, stating that a block trade is executed at an average price per share that is somewhere between the asset prices observed immediately before and after the execution. The more general case $c \geq 0$ is just technical at this stage but will be needed in Section 2.4.4. Assumption (2.7) means that proceeds by a simple strategy as in (2.9) are

$$L_t(\Theta^n) = -\sum_{k:\ t_k \leq t} \xi_k (\Theta_{t_k} - \Theta_{t_{k-1}}) \tag{2.8}$$

for some random variable ξ_k between $g(\overline{S}_{t_k}, Y^{\Theta^n}_{t_k-})$ and $g(\overline{S}_{t_k}, Y^{\Theta^n}_{t_k-} + c\Delta Y^{\Theta^n}_t)$. Note that at this point we have not specified the proceeds (negative expenses) from block trades, but we only assume that they satisfy some natural bounds. Yet, this is indeed already sufficient to derive the functional (2.4) for continuous strategies as a limit of simple ones.

Lemma 2.2.1. *For $T > 0$, approximate a continuous f.v. process $(\Theta_t)_{t\in[0,T]}$ by a sequence $(\Theta^n_t)_{t\in[0,T]}$ of simple trading strategies given as follows: For a sequence of partitions $\{0 = t_0 < t_1 < \cdots < t_{m_n} = T\}$, $n \in \mathbb{N}$, with $\sup_{1\leq k\leq m_n}|t_k - t_{k-1}| \to 0$ for*

$n \to \infty$, let

$$\Theta_t^n := \Theta_0 + \sum_{k=1}^{m_n} \big(\Theta_{t_k} - \Theta_{t_{k-1}}\big) \mathbb{1}_{[t_k,T]}(t)\,, \quad t \in [0,T]. \tag{2.9}$$

Assume (2.7) *holds for some $c \geq 0$. Then* $\sup_{0\leq t\leq T}|L_t(\Theta^n) + \int_0^t S_u \,\mathrm{d}\Theta_u| \xrightarrow{n\to\infty} 0$ *a.s.*

Proof. Note that $\sup_{u\in[0,T]}|\Theta_u^n - \Theta_u| \to 0$ as $n \to \infty$. The solution map $\Theta \mapsto Y^\Theta$ is continuous with respect to the uniform norm, see Proposition 2.5.1. Therefore,

$$\sup_{u\in[0,T]} |Y_u^{\Theta^n} - Y_u^\Theta| \to 0 \quad \text{a.s. for } n \to \infty. \tag{2.10}$$

Note that for $\Delta\Theta_{t_k} := \Theta_{t_k} - \Theta_{t_{k-1}}$ and ξ_k between $g(\overline{S}_{t_k}, Y_{t_k-}^{\Theta^n})$ and $g(\overline{S}_{t_k}, Y_{t_k-}^{\Theta^n} + c\Delta\Theta_{t_k})$ and $Y := Y^\Theta$ we have

$$\begin{aligned} |\xi_k - g(\overline{S}_{t_k}, Y_{t_k})| &\leq L_g(\overline{S}_{t_k}, \omega) \max\big\{\big|Y_{t_k} - Y_{t_k-}^{\Theta^n} - c\Delta\Theta_{t_k}\big|, \big|Y_{t_k} - Y_{t_k-}^{\Theta^n}\big|\big\} \\ &\leq \tilde{c} L_g(\overline{S}_{t_k}, \omega)\big(\big|Y_{t_k} - Y_{t_k}^{\Theta^n}\big| + |\Delta\Theta_{t_k}|\big), \end{aligned}$$

where $\tilde{c} > 0$ is a universal constant, $L_g(x,\omega)$ denotes the Lipschitz constant of $y \mapsto g(x,y)$ on a compact set, depending on the (bounded) realizations for $\omega \in \Omega$ of Y^Θ and Y^{Θ^n}, $n \in \mathbb{N}$, on the interval $[0,T]$; such a compact set exits since Θ is continuous and $\sup_{u\in[0,T]}\big|Y_u^\Theta - Y_u^{\Theta^n}\big|$ can be bounded by a factor times the uniform distance between Θ and Θ^n on $[0,T]$, cf. [PTW07, proof of Thm. 4.1]. Hence,

$$L_t(\Theta^n) = -\sum_{k:\ t_k\leq t} g(\overline{S}_{t_k}, Y_{t_k}^\Theta)\big(\Theta_{t_k} - \Theta_{t_{k-1}}\big) + \mathcal{E}_t^n\,, \tag{2.11}$$

$$\text{where } |\mathcal{E}_t^n| \leq \tilde{c}\Big(\sup_{u\in[0,T]} L_g(\overline{S}_u, \omega)\Big) \sum_{k=1}^{m_n} \big(\big|Y_{t_k} - Y_{t_k}^{\Theta^n}\big| + |\Delta\Theta_{t_k}|\big)|\Delta\Theta_{t_k}| \tag{2.12}$$

$$\leq C(\omega)\Big(\sup_{1\leq k\leq m_n} \big|Y_{t_k} - Y_{t_k}^{\Theta^n}\big|\Big)|\Theta(\omega)|_{\mathrm{TV}} + C(\omega)\sum_{k=1}^{m_n}|\Delta\Theta_{t_k}|^2 \tag{2.13}$$

$$\to 0 \quad \text{a.s. for } n \to \infty \text{ (uniformly in } t\text{)}, \tag{2.14}$$

thanks to (2.10) and the fact that Θ has continuous paths of finite variation. The claim follows since by dominated convergence the Riemann-sum process in (2.11) converges a.s. to the Stieltjes-integral process $-\int_0^\cdot S_u \,\mathrm{d}\Theta_u$ uniformly on $[0,T]$. □

Example 2.2.2 (Continuity issues for an alternative "ad-hoc" definition of proceeds)**.** Consider the problem of optimally liquidating $\Theta_{0-} = 1$ risky asset in time $[0,T]$ while maximizing expected proceeds. In view of assumption (2.7), an alternative but possibly "ad-hoc" definition for proceeds $\tilde{L}_T$ of simple strategies could be to consider just some price for each block trade, similarly to [BB04, Section 3] or [HH11, Example 2.4]. For multiplicative impact $g(\overline{S}, Y) = \overline{S} f(Y)$, taking e.g. the price directly after the impact would yield for simple strategies Θ^n that trade at times $\{0 = t_0^n < t_1^n < \cdots < t_n^n = T\}$

the proceeds $\tilde{L}_T(\Theta^n) = -\sum_{k=0}^n \overline{S}_{t_k^n} f(Y_{t_k^n}^{\Theta^n})\Delta\Theta_{t_k^n}^n$. The family $(\Theta^n)_n$ of strategies which liquidate an initial position of size 1 until time $1/n$ in n equidistant blocks of uniform size is given by $\Theta_t^n := \sum_{k=1}^n \frac{n-k+1}{n}\mathbb{1}_{[\frac{k-1}{n^2},\frac{k}{n^2})}(t)$. With unaffected price $\overline{S}_t = e^{-\delta t}\widetilde{M}_t$ for a continuous martingale $\widetilde{M}$, and permanent impact ($h \equiv 0$), i.e. $Y_t = \Theta_t - 1$, this yields $\mathbb{E}[\tilde{L}_T(\Theta^n)] \to \int_0^1 f(-y)\,\mathrm{d}y$ for $n \to \infty$. Given $\delta \geq 0$, for any non-increasing simple strategy $\Theta = \sum_{k=1}^n \Theta_{\tau_k}\mathbb{1}_{[\![\tau_{k-1},\tau_k[\![}$ with $\Theta_{0-} = 1$ holds that $\mathbb{E}[\tilde{L}(\Theta)] \leq \int_0^1 f(-y)\,\mathrm{d}y$ with strict inequality for $\delta > 0$. So the control sequence (Θ^n) is only asymptotically optimal among all simple monotone liquidation strategies.

Remark 2.2.3. Note that Example 2.2.2 is a toy example, since for permanent impact the optimal strategy (considering asymptotically realizable proceeds) is trivial and in case $\delta = 0$ any strategy is optimal, cf. [GZ15, Prop. 3.5(III) and the comment preceding it]. Nevertheless, this example shows that the object of interest are *asymptotically realizable* proceeds, an insight due to [BB04]. For analysis, it thus appears convenient and sensible not to make a formal distinction of (sub-optimal) realizable and asymptotically realizable proceeds, but to consider the latter and interpret strategies accordingly. Investigating asymptotically realizable proceeds can help to answer questions on modeling issues, e.g. whether the large investor could sidestep liquidity costs entirely and in effect act as a small investor, cf. [BB04, ÇJP04]. One could impose, like [ÇST10], additional constraints on strategies to avoid such issues; But in such tweaked models one could not investigate the effects from some given illiquidity friction alone, in isolation from other constraints, because results from an analysis will be consequences of the combination of both frictions.

Using integration-by-parts, we can obtain the following alternative representation of the functional in (2.4) for continuous f.v. strategies:

$$\begin{aligned}
L(\Theta) = &\int_0^{\cdot} G_x(\overline{S}_{u-}, Y_{u-}^{\Theta})\,\mathrm{d}\overline{S}_u + \int_0^{\cdot} \Big(\tfrac{1}{2}G_{xx}(\overline{S}_u, Y_u^{\Theta})\overline{S}_u^2 - g(\overline{S}_u, Y_u^{\Theta})h(Y_u^{\Theta})\Big)\,\mathrm{d}\langle M\rangle_u \\
&- \big(G(\overline{S}_\cdot, Y_\cdot^{\Theta}) - G(\overline{S}_0, Y_{0-}^{\Theta})\big) \\
&+ \sum_{\substack{\Delta\overline{S}_u \neq 0 \\ 0\leq u\leq \cdot}} \big(G(\overline{S}_u, Y_u^{\Theta}) - G(\overline{S}_{u-}, Y_u^{\Theta}) - G_x(\overline{S}_{u-}, Y_u^{\Theta})\Delta\overline{S}_u\big)\,, \qquad (2.15)
\end{aligned}$$

where $G(x,y) := \int_c^y g(x,z)\,\mathrm{d}z$ for constant c, and using that $\overline{S}$ and Y have no common jumps. The advantage of this representation is that the right-hand side of (2.15) makes sense for any predictable process Θ with càdlàg paths in contrast to the term in (2.4) This form of the proceeds will turn out to be helpful for the stability analysis. We will show that the right-hand side in (2.15) is continuous in the control Θ when the path-space of Θ, the càdlàg path space, is endowed with various topologies. Hence, it can be used to define the proceeds for general trading strategies by continuity. Next section is going to discuss the topologies that will be of interest.

2.2.1 The Skorokhod space and its M_1 and J_1 topologies

We are going to derive a continuity result (Theorem 2.2.7) for the functional L in different topologies on the space $D \equiv D([0,T]) := D([0,T];\mathbb{R})$ of real-valued càdlàg paths on the time interval $[0,T]$. Following the convention by [Sko56], we take each element in $D[0,T]$ to be left-continuous at time T.[1] One could also consider initial and terminal jumps by extending the paths, see Remark 2.2.6. At this point, let us remark that finite horizon T is not essential for the results below, whose analysis carries over to the time interval $[0,\infty)$ because the topology on $D([0,\infty))$ is induced by the topologies of $D([0,T])$ for $T \geq 0$. More precisely, for the topologies we are interested in, $x_n \to x$ as $n \to \infty$ in $D([0,\infty))$ if $x_n \to x$ in $D([0,t])$ for the restrictions of x_n, x on $[0,t]$, for any t being a continuity point of x, see [Whi02, Sect. 12.9].

Convergence in the uniform topology is rather strong, in that approximating a path with a jump is only possible if the approximating sequence has jumps of comparable size at the same time. If one is interested in stability with respect to slight shift of the execution in time, then a familiar choice that also makes D separable, the Skorokhod J_1 topology, might be appropriate; for comprehensive study, see [Bil99, Ch. 3]. However, also here an approximating sequence for a path with jumps needs jumps of comparable size, if only at nearby times. To capture the occurrence of the so-called *unmatched jumps*, i.e. jumps that appear in the limit of continuous processes, another topology on D is more appropriate, the Skorokhod M_1 topology. Recall that $x_n \to x$ in (D, d_{M_1}) if $d_{M_1}(x_n, x) \to 0$ as $n \to \infty$, with

$$d_{M_1}(x_n, x) := \inf\big\{ \|u - u_n\| \vee \|r - r_n\| \;\big|\; (u,r) \in \Pi(x), (u_n, r_n) \in \Pi(x_n) \big\}\,, \quad (2.16)$$

where $\|\cdot\|$ denotes the uniform norm on $[0,1]$ and $\Pi(x)$ is the set of all *parametric representations* $(u,r) : [0,1] \to \Gamma(x)$ of the completed graph (with vertical connections at jumps) $\Gamma(x)$ of $x \in D$, see [Whi02, Sect. 3.3]. In essence, two functions $x, y \in D$ are near to each other in M_1 if one could run continuously a particle on each graph $\Gamma(x)$ and $\Gamma(y)$ from the left endpoint toward the right endpoint such that the two particles are nearby in time and space. In particular, it is easy to see that a simple jump path could be approximated in M_1 by a sequence of absolutely continuous paths, in contrast to the uniform and the J_1 topologies. More precisely, we have the following

Proposition 2.2.4. *Let $x \in D([0,T])$ and consider the Wong-Zakai-type approximation sequence $(x_n) \subset D([0,T])$ defined by $x_n(t) := n \int_{t-1/n}^{t} x(s)\,\mathrm{d}s$, $t \in [0,T]$. Then*

$$x_n \to x \quad \textit{for } n \to \infty, \quad \textit{in } (D([0,T]), M_1).$$

Proof. To ease notation, we embed a path x in $D([0,\infty))$ and consider the corresponding approximating sequence for the extended path on $[0,\infty)$. The claim follows by restricting to the domain $[0,T]$, as 0 and T are continuity points of x, cf. [Whi02, Sect. 12.9]. The idea is to construct explicitly parametric representations of $\Gamma(x)$ and $\Gamma(x_n)$ that

[1] This is implicitly assumed also in [Whi02], see the compactness criterion in Thm. 12.12.2 which is borrowed from [Sko56].

are close enough. For this purpose, we need to add "fictitious" time to be able to parametrize the segments that connect jump points of x. Indeed, let (a_k) be a fixed convergent series of strictly positive numbers and let $t_1, t_2, \dots$ be the jump times of x ordered such that $|\Delta x(t_1)| \geq |\Delta x(t_2)| \geq \dots$ and $t_k < t_{k+1}$ if $|\Delta x(t_k)| = |\Delta x(t_{k+1})|$. Set $\delta(t) := \sum_k a_k \mathbb{1}_{\{t_k \leq t\}}$, the total "fictitious" time added to parametrize the jumps of x up to time t.

Consider the time-changes $\gamma_n(t) := n \int_{t-1/n}^{t} (\delta(u) + u)\,\mathrm{d}u$ and $\gamma_0(t) := \delta(t) + t$, $t \geq 0$, together with their continuous inverses $\gamma_n^{-1}(s) := \inf\{u > 0 \mid \gamma_n(u) > s\}$ for $s \geq 0$, $n \geq 0$. It is easy to check that we have

$$\gamma_n^{-1}(s) - 1/n < \gamma_0^{-1}(s) < \gamma_n^{-1}(s) < \infty \quad \text{for } s \geq 0, \tag{2.17}$$

because $\gamma_n(t) < \gamma_0(t) < \gamma_n(t + 1/n)$, cf. [KPP95, Lemma 6.1]. Consider the sequence $u_n(s) := x_n(\gamma_n^{-1}(s))$ for $s \geq 0$ and let

$$u(s) := \begin{cases} x(\gamma_0^{-1}(s)) & \text{if } \eta_1(s) = \eta_2(s), \\ x(\gamma_0^{-1}(s)) \cdot \frac{s-\eta_1(s)}{\eta_2(s)-\eta_1(s)} + x(\gamma_0^{-1}(s)-) \cdot \frac{\eta_2(s)-s}{\eta_2(s)-\eta_1(s)} & \text{if } \eta_1(s) \neq \eta_2(s), \end{cases}$$

where $[\eta_1(s), \eta_2(s)]$ is the "fictitious" time added for a jump at time $t = \gamma_0^{-1}(s)$, i.e. $\eta_1(s) := \sup\{\tilde{s} \mid \gamma_0^{-1}(\tilde{s}) < \gamma_0^{-1}(s)\}$ and $\eta_2(s) := \inf\{\tilde{s} \mid \gamma_0^{-1}(\tilde{s}) > \gamma_0^{-1}(s)\}$, as in [KPP95, p. 368]. Then [KPP95, Lemma 6.2] gives $\lim_{n\to\infty} u_n = u$, uniformly on bounded intervals; our setup corresponds to $f \equiv 1$ there, so our u_n, u correspond to $V^{1/n}, V$ there.

Now the claim follows by observing that (u_n, γ_n^{-1}) is a parametric representation of the completed graph of x_n, i.e. $(u_n, \gamma_n^{-1}) \in \Pi(x_n)$, and $(u, \gamma_0^{-1}) \in \Pi(x)$ which are arbitrarily close when n is big. □

Remark 2.2.5. A direct corollary of Proposition 2.2.4 is that $D([0,T])$ is the closure of the set of absolutely continuous functions in the Skorokhod M_1 topology, in contrast to the uniform or Skorokhod J_1 topologies where a jump in the limit can only be approximated by jumps of comparable sizes.

Remark 2.2.6 (Extended paths)**.** To include trading strategies that could additionally have initial and terminal jumps in our analysis, one may embed the paths of such strategies in the slightly larger space $D([-\varepsilon, T+\varepsilon]; \mathbb{R})$ for some $\varepsilon > 0$, e.g. $\varepsilon = 1$, by setting $x(s) = x(0-)$ for $s \in [-\varepsilon, 0)$ and $x(s) = x(T+)$ for $s \in (T, T+\varepsilon]$; we will refer to thereby embedded paths as *extended paths*. This extension is relevant when trying to approximate jumps at terminal time by absolutely continuous strategies in a non-anticipative way as e.g. in Proposition 2.2.4 where it is clear that a bit more time could be required after a jump occurs in order to approximate it. In particular, by considering extended paths the result of Proposition 2.2.4 holds if one allows for initial and terminal jumps of x, but convergence holds in the extended paths space.

2.2.2 Main stability results

Our main result is stability of the functional L defined by the right-hand side of (2.15) for processes Θ with càdlàg paths.

Theorem 2.2.7. *Let a sequence of predictable processes (Θ^n) converge to the predictable process Θ in (D, ρ), in probability, where ρ denotes the uniform topology, the Skorokhod J_1 or M_1 topology, being generated by a suitable metric d. Assume that (Θ^n) is bounded in $L^0(\mathbb{P})$, i.e. there exists $K \in L^0(\mathbb{P})$ such that $\sup_{0\le t\le T} |\Theta^n_t| \le K$ for all n. Then the sequence of processes $L(\Theta^n)$ converges to $L(\Theta)$ in (D, ρ) in probability, i.e.*

$$\mathbb{P}\big[d\big(L(\Theta^n), L(\Theta)\big) \ge \varepsilon\big] \to 0 \qquad \text{for } n \to \infty \text{ and } \varepsilon > 0. \tag{2.18}$$

In particular, there is a subsequence $L(\Theta^{n_k})$ that converges a.s. to $L(\Theta)$ in (D, ρ).

Note that e.g. for almost sure convergence $\Theta^n \to \Theta$ in (D, ρ), the $L^0(\mathbb{P})$ boundedness condition is automatically fulfilled.

Proof. By considering subsequences, one could assume that the sequence (Θ^n) converges to Θ in (D, ρ) a.s. The idea for the proof is to show that each summand in the definition of L is continuous. But as D endowed with J_1 or M_1 is not a topological vector space, since addition is not continuous in general, further arguments will be required. Addition is continuous (and hence also multiplication) if for instance the summands have no common jumps, see [JS03, Prop. VI.2.2] for J_1 and [Whi02, Cor. 12.7.1] for M_1. In our case however, there are three terms in L that can have common jumps, namely the stochastic integral process $\int_0^\cdot G_x(\overline{S}_{u-}, Y_{u-})\,\mathrm{d}\overline{S}_u$, the sum $\Sigma := \sum_{u\le\cdot}\big(G(\overline{S}_u, Y_u) - G(\overline{S}_{u-}, Y_u) - G_x(\overline{S}_{u-}, Y_u)\Delta\overline{S}_u\big)$ of jumps and the term $-G(\overline{S}, Y)$. At jump times of Θ (i.e. of Y) which are predictable stopping times, $\overline{S}$ does not jump since it is quasi-left continuous. Hence the only common jump times can be jumps times of $\overline{S}$ which are totally inaccessible. If $\Delta\overline{S}_\tau \neq 0$, we have then $\Delta(\int_0^\cdot G_x(\overline{S}_{u-}, Y_{u-})\,\mathrm{d}\overline{S}_u)_\tau = G_x(\overline{S}_{\tau-}, Y_\tau)\Delta\overline{S}_\tau$ and also $\Delta(-G(\overline{S}, Y))_\tau = -\big(G(\overline{S}_\tau, Y_\tau) - G(\overline{S}_{\tau-}, Y_\tau)\big)$, because $\Delta Y_\tau = 0$ a.s. Since moreover

$$\Delta\Sigma_\tau = G(\overline{S}_\tau, Y_\tau) - G(\overline{S}_{\tau-}, Y_\tau) - G_x(\overline{S}_{\tau-}, Y_\tau)\Delta\overline{S}_\tau,$$

one has cancellation of jumps at jump times of $\overline{S}$. However, these are times of continuity for Y and this will be crucial below to deduce continuity of addition on the support of $\big(\int_0^\cdot G_x(\overline{S}_{u-}, Y_{u-})\,\mathrm{d}\overline{S}_u, \Sigma, -G(\overline{S}, Y)\big)$ in $(D, \rho) \times (D, \rho) \times (D, \rho)$.

First consider the case of uniformly bounded sequence (Θ^n). Then the processes

$$\mathrm{d}Y^n_t = -h(Y^n_t)\,\mathrm{d}\langle M\rangle_t + \mathrm{d}\Theta^n_t, \quad Y^n_{0-} = y\,,$$

are uniformly bounded, so we can assume w.l.o.g. that h, gh, G, G_x and G_{xx} are ω-wise Lipschitz continuous and bounded (it is so on the range of all Y^n, Y, which is contained in a compact subset of $\mathbb{R}$). By Proposition 2.5.1 we have $Y^n \to Y$ in (D, ρ), almost surely. This implies $(\overline{S}, Y^n) \to (\overline{S}, Y)$ almost surely, by absence of common jumps of $\overline{S}$

and Y, cf. [JS03, Prop. VI.2.2b] for J_1 and[2] [Whi02, Thm. 12.6.1 and 12.7.1] for M_1. By the Lipschitz property of G and (for the M_1 case) monotonicity of $G(\cdot, y)$ and $G(x, \cdot)$, we get

$$G(\overline{S}, Y^n) \to G(\overline{S}, Y) \quad \text{in } (D, \rho), \text{ a.s.} \tag{2.19}$$

Indeed, for the M_1 topology, it is easy to see that $(G(u^1, u^2), r) \in \Pi(G(\overline{S}, Y))$ for any parametric representation $((u^1, u^2), r)$ of $(\overline{S}, Y)$, because at jump times t of $G(\overline{S}, Y)$, $z \mapsto r(z) \equiv t$ is constant on an interval $[z_1, z_2]$, and either u^1 or u^2 is constant on $[z_1, z_2]$.

Note that jump times of Θ and Y coincide, and form a random countable subset of $[0, T]$. Moreover, convergence in (D, ρ) implies local uniform convergence at continuity points of the limit (for ρ being the M_1 topology, cf. [Whi02, Lemma 12.5.1], for the J_1 topology cf. [JS03, Prop. VI.2.1]). Hence, $Y_t^n \to Y_t$ for almost all $t \in [0, T]$, $\mathbb{P}$-a.s. By Lipschitz continuity of G_{xx} and gh, we get $\frac{1}{2}G_{xx}(\overline{S}_t, Y_t^n) - g(\overline{S}_t, Y_t^n)h(Y_t^n) \to \frac{1}{2}G_{xx}(\overline{S}_t, Y_t) - g(\overline{S}_t, Y_t)h(Y_t)$, for almost-all $t \in [0, T]$, $\mathbb{P}$-a.s. By dominated convergence, we conclude that

$$\int_0^\cdot \big(\tfrac{1}{2}G_{xx}(\overline{S}_u, Y_u^n) - g(\overline{S}_u, Y_u^n)h(Y_u^n)\big) \, \mathrm{d}\langle M\rangle_u \to \int_0^\cdot \big(\tfrac{1}{2}G_{xx}(\overline{S}_u, Y_u) - g(\overline{S}_u, Y_u)h(Y_u)\big) \, \mathrm{d}\langle M\rangle_u$$

uniformly on $[0, T]$, a.s., using that $\langle M\rangle$ is absolutely continuous w.r.t. Lebesgue measure. Hence these two summands in the definition of L, see (2.15), are (ω-wise) continuous in Θ.

Now we treat the stochastic integral and jump terms in (2.15). By the above arguments we can also deal with the drift in the process $\overline{S}$. Thus we may assume w.l.o.g. that $\overline{S}$ is a martingale. In particular, up to a localization argument (see below for details), we can assume that $\overline{S}$ is bounded and therefore the stochastic integral is a true martingale, since the integrand is bounded. Having $Y^n \to Y$ a.e. on the space $(\Omega \times [0, T], \mathbb{P} \otimes \mathrm{Leb}([0, T]))$, we can conclude convergence of the stochastic integrals in the uniform topology, in probability. Dominated convergence on $\big([0, T], \mathrm{Leb}([0, T])\big)$ yields

$$\int_0^T (Y_{u-}^n - Y_{u-})^2 \, \mathrm{d}\langle \overline{S}\rangle_u \to 0 \quad \text{as } n \to \infty, \ \mathbb{P}\text{-a.s.}$$

Since Y^n, Y are uniformly bounded one gets, again by dominated convergence, that

$$\mathbb{E}\Big[\int_0^T (Y_{u-}^n - Y_{u-})^2 \, \mathrm{d}\langle \overline{S}\rangle_u\Big] \to 0 \quad \text{as } n \to \infty,$$

i.e. $Y_-^n \to Y_-$ in $L^2(\Omega \times [0, T], \mathrm{d}\mathbb{P} \otimes \mathrm{d}\langle \overline{S}\rangle)$. By localization (to bound $\overline{S}$ and use that $G_x(x, y)$ is locally Lipschitz in y), Itô's isometry and Doob's inequality, we get

$$\mathbb{P}\Big[\sup_{0 \le t \le T}\Big|\int_0^t G_x(\overline{S}_{u-}, Y_{u-}^n) \, \mathrm{d}\overline{S}_u - \int_0^t G_x(\overline{S}_{u-}, Y_{u-}) \, \mathrm{d}\overline{S}_u\Big| \ge \varepsilon\Big] \to 0 \quad \text{as } n \to \infty. \tag{2.20}$$

[2]Using the strong M_1 topology in $D([0, \infty); \mathbb{R}^2)$.

For the sum of jumps Σ^n (defined like Σ, but with Y^n instead of Y) we have a.s. uniform convergence $\Sigma^n \to \Sigma$ by Lemma 2.5.4. Hence $\int_0^t G_x(\overline{S}_{u-}, Y^n_{u-})\,\mathrm{d}\overline{S}_u + \Sigma^n$ converges in ucp. To conclude on the proceeds, note that at jump times of $\overline{S}$, when cancellation of jumps occurs, one has continuity of Y and hence local uniform convergence of the sequence Y^n. For our setup, Lemmas 2.5.2 and 2.5.3 show continuity of addition on the support of $\big(\int_0^\cdot G_x(\overline{S}_{u-}, Y_{u-})\,\mathrm{d}\overline{S}_u + \Sigma, -G(\overline{S}, Y)\big)$ (along the support of $\big(\int_0^\cdot G_x(\overline{S}_{u-}, Y^n_{u-})\,\mathrm{d}\overline{S}_u + \Sigma^n, -G(\overline{S}, Y^n)\big)\big)$ for the J_1 and M_1 topologies, respectively. So the continuous mapping theorem [Kal02, Lem. 4.3] yields the claim for the proceeds functional L (the uniform topology being stronger than ρ).

It remains to investigate the more general case of $\overline{S}$ and (Θ^n) being only bounded in $L^0(\mathbb{P})$. Note that the continuity of all terms except the stochastic integral in the definition of L was proven ω-wise; in this case $\sup_n \sup_{0\le t\le T}|\Theta^n_t(\omega)| < \infty$ (by the a.s. convergence of Θ^n to Θ in (D,ρ)) and hence the same arguments carry over here by restricting our attention to compact sets (depending on ω). Hence refinement of the argument above is only needed for the stochastic integral term. The bound on $\overline{S}$ and (Θ^n) means that for every $\varepsilon > 0$ there exists $\Omega_\varepsilon \in \mathcal{F}$ with $\mathbb{P}(\Omega_\varepsilon) > 1-\varepsilon$ and a positive constant K_ε which is a uniform bound for the sequence (together with the limit Θ) on Ω_ε. For the stopping time $\tau := \inf \tau_n$, where $\tau_n := \inf\{t \ge 0 \mid |\Theta^n_t| \vee |\overline{S}_t| > K_\varepsilon\} \wedge T$ (τ is a stopping time because the filtration is right-continuous by our assumptions), we then have that $\tau = T$ on Ω_ε. By the arguments above we conclude that $d\big(\int_0^{\cdot\wedge\tau} G_x(\overline{S}_{u-}, Y^n_{u-})\,\mathrm{d}\overline{S}_u, \int_0^{\cdot\wedge\tau} G_x(\overline{S}_{u-}, Y_{u-})\,\mathrm{d}\overline{S}_u\big) \to 0$ in probability. Since $\int_0^{\cdot\wedge\tau} G_x(\overline{S}_{u-}, Y^n_{u-})\,\mathrm{d}\overline{S}_u = \int_0^\cdot G_x(\overline{S}_{u-}, Y^n_{u-})\,\mathrm{d}\overline{S}_u$ on Ω_ε, we conclude

$$\mathbb{P}\left[d\Big(\int_0^\cdot G_x(\overline{S}_{u-}, Y^n_{u-})\,\mathrm{d}\overline{S}_u, \int_0^\cdot G_x(\overline{S}_{u-}, Y_{u-})\,\mathrm{d}\overline{S}_u\Big) \ge \varepsilon\right] \le 2\varepsilon$$

for all n large enough, and this finishes the proof since ε was arbitrary. □

Remark 2.2.8. Inspection of the proof above reveals that predictability of the strategies is only needed to show why the addition map is continuous when there is cancellation of jumps in (2.15); indeed, for predictable Θ the processes Y^Θ and $\overline{S}$ will have no common jump and this was sufficient for the arguments. However, in the case when M (and thus $\overline{S}$) is continuous, only one term in (2.15) might have jumps, namely $G(\overline{S}, Y^\Theta)$. Hence, in this case the conclusion of Theorem 2.2.7 even holds under the relaxed assumption that the càdlàg strategies are merely adapted, instead of being predictable.

Remark 2.2.9. Our assumption of positive prices (and monotonicity of $x \mapsto g(x,y)$) has been (just) used to prove the M_1-convergence of $G(\overline{S}, Y^n)$ in (2.19). If one would want to consider a model where prices could become negative (like additive impact $S = \overline{S} + f(Y)$, see Example 2.1.1), then M_1-continuity of proceeds would not hold in general, as a simple counter-example can show. Yet, the above proof still shows $L_t(\Theta^n) \to L_t(\Theta)$ in probability, for all $t \in [0,T]$ where $\Delta\Theta_t = 0$. Also note that for continuous Θ^n converging in M_1 to a continuous strategy Θ, hence also uniformly, one obtains that proceeds $L(\Theta^n) \to L(\Theta)$ converge uniformly, in probability.

An important consequence of Theorem 2.2.7 is a stability property for our model. It essentially implies that we can approximate each strategy by a sequence of absolutely continuous strategies, corresponding to small intertemporal shifts of reassigned trades, whose proceeds will approximate the proceeds of the original strategy. More precisely, if we restrict our attention to the class of monotone strategies, then we can restate this stability in terms of the Prokhorov metric on the pathwise proceeds (which are monotone and hence define measures on the time axis). This result on stability of proceeds with respect to small intertemporal Wong-Zakai-type re-allocation of orders may be compared to seminal work by [HHK92] on a different but related problem, who required that for economic reason the utility should be a continuous functional of cumulative consumption with respect to the Lévy–Prokhorov metric d_{LP}, in order to satisfy the sensible property of intertemporal substitution for consumption. Recall for convenience of the reader the definition of d_{LP} in our context: for increasing càdlàg paths on $[0,\tilde{T}]$, $x, y : [0,\tilde{T}] \to \mathbb{R}$ with $x(0-) = y(0-)$ and $x(\tilde{T}) = y(\tilde{T})$,

$$d_{\mathrm{LP}}(x,y) := \inf\{\varepsilon > 0 \mid x(t) \leq y((t+\varepsilon)\wedge\tilde{T})+\varepsilon,\ \ y(t) \leq x((t+\varepsilon)\wedge\tilde{T})+\varepsilon \ \ \forall t \in [0,\tilde{T}]\}.$$

Corollary 2.2.10. *Let Θ be a predictable process with càdlàg paths defined on the time interval $[0,T]$ (with possible initial and terminal jumps) that is extended to the time interval $[-1, T+1]$ as in Remark 2.2.6. Consider the sequence of f.v. processes (Θ^n) where*

$$\Theta^n_t := n \int_{t-1/n}^{t} \Theta_s \,\mathrm{d}s, \quad t \geq 0, \tag{2.21}$$

and let $L := L(\Theta), L^n := L(\Theta^n)$ be the proceeds processes from the respective trading. Then $L^n_t \to L_t$ at all continuity points $t \in [0, T+1]$ of L as $n \to \infty$, in probability. In particular, for any bounded monotone strategy Θ the Borel measures $L^n(\mathrm{d}t;\omega)$ and $L(\mathrm{d}t;\omega)$ on $[0,T+1]$ are finite (a.s.) and converge in the Lévy–Prokhorov metric $d_{LP}(L^n(\omega), L(\omega))$ in probability, i.e. for any $\varepsilon > 0$,

$$\mathbb{P}\big[d_{LP}(L^n(\omega), L(\omega)) > \varepsilon\big] \to 0 \quad \text{as } n \to \infty.$$

Proof. An application of Proposition 2.2.4 together with Theorem 2.2.7 gives

$$d_{M1}(L^n, L) \xrightarrow{\mathbb{P}} 0.$$

The first part of the claim now follows from the fact that convergence in M_1 implies local uniform convergence at continuity points of the limit, see [Whi02, Lemma 12.5.1]. The same property implies the claim about the Lévy–Prokhorov metric because convergence in this metric is equivalent to weak convergence of the associated measures which on the other hand is equivalent to convergence at all continuity points of the cumulative distribution function (together with the total mass). □

Note that the sequence (Θ^n) from Corollary 2.2.10 satisfies $\Theta^n \equiv \Theta_T$ on $[T+1/n, T+1]$

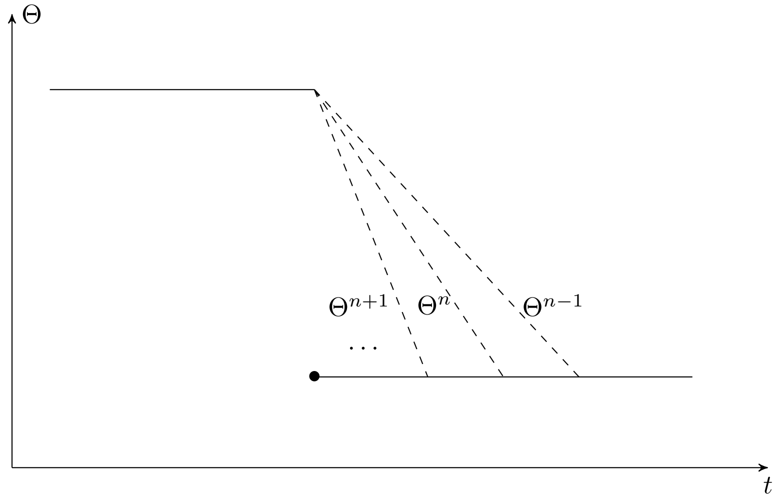


Figure 2.1: The Wong–Zakai approximation in (2.21) for a single jump process.

for all n, i.e. the approximating strategies arrive at the position Θ_T, however by requiring a bit more time to execute. Based on the Wong–Zakai approximation sequence from (2.21), we next show that each semimartingale strategy on the time interval $[0,T]$ can be approximated by simple adapted strategies with uniformly small jumps that, however, again need slightly more time to be executed.

Proposition 2.2.11. *Let $(\Theta_t)_{t\in[0,T]}$ be a predictable process with càdlàg paths extended to the time interval $[0,T+1]$ as in Remark 2.2.6. Then there exists a sequence $(\Theta^n_t)_{t\in[0,T+1]}$ of simple predictable càdlàg processes with jumps of size not more than $1/n$ such that $d_{M1}(L(\Theta^n),L(\Theta)) \xrightarrow{\mathbb{P}} 0$ as $n\to\infty$, where d_{M1} denotes the Skorokhod M_1 metric on $D([0,T+1];\mathbb{R})$. Moreover, if Θ is continuous, the same convergence holds true in the uniform metric on $[0,T]$ instead.*

Proof. Consider the Wong-Zakai-type approximation sequence $\widetilde{\Theta}^n$ from Corollary 2.2.10 for which $d_{M_1}(L(\widetilde{\Theta}^n),L(\Theta)) \xrightarrow{\mathbb{P}} 0$, where the Skorokhod M_1 topology is considered for the extended paths on time-horizon $[0,T+1]$. Now we approximate each (absolutely) continuous process $\widetilde{\Theta}^n$ by a sequence of simple processes as follows.

For $\varepsilon>0$, consider the sequence of stopping times with $\sigma_0^{\varepsilon,n} := 0$ and

$$\sigma_{k+1}^{\varepsilon,n} := \inf\big\{t \;\big|\; t>\sigma_k^{\varepsilon,n} \text{ and } |\widetilde{\Theta}^n_t - \widetilde{\Theta}^n_{\sigma_k^{\varepsilon,n}}| \geq \varepsilon\big\} \wedge (\sigma_k^{\varepsilon,n}+1/n) \quad \text{for } k\geq 0.$$

Note that $\sigma_k^{\varepsilon,n}$ are predictable as hitting times of continuous processes and $\sigma_k^{\varepsilon,n} \nearrow \infty$ as $k\to\infty$ because the process $\widetilde{\Theta}^n$ is continuous. When $\varepsilon\to 0$, we have $\Theta^{\varepsilon,n} \xrightarrow{ucp} \widetilde{\Theta}^n$ for

$$\Theta^{\varepsilon,n} := \widetilde{\Theta}^n_0 + \sum_{k=1}^{\infty}\big(\widetilde{\Theta}^n_{\sigma_k^{\varepsilon,n}} - \widetilde{\Theta}^n_{\sigma_{k-1}^{\varepsilon,n}}\big)\mathbb{1}_{[\![\sigma_k^{\varepsilon,n},\infty[\![}.$$

Moreover, if for each integer $m \geq 1$ we define the (predictable) process $\Theta^{\varepsilon,n,m}$ by

$$\Theta^{\varepsilon,n,m} := \widetilde{\Theta}_0^n + \sum_{k=1}^{m} \big(\widetilde{\Theta}^n_{\sigma_k^{\varepsilon,n}} - \widetilde{\Theta}^n_{\sigma_{k-1}^{\varepsilon,n}}\big) \mathbb{1}_{[\![\sigma_k^{\varepsilon,n},\infty[\![},$$

then for each fixed ε and n we have $\Theta^{\varepsilon,n,m} \xrightarrow{ucp} \Theta^{\varepsilon,n}$ when $m \to \infty$. Hence, we can choose $\varepsilon = \varepsilon(n)$ small enough and $m = m(n)$ big enough such that

$$d(\widetilde{\Theta}^n, \Theta^{\varepsilon(n),n,m(n)}) < 2^{-n},$$

with $d(\cdot,\cdot)$ denoting a metric that metrizes ucp convergence (cf. e.g. [Pro05, p. 57]). Thus, $\Theta^n := \Theta^{\varepsilon(n),n,m(n)}$ will be close to Θ in the Skorokhod M_1 topology, in probability, because the uniform topology is stronger than the M_1 topology.

Note that if Θ is already continuous, no intermediate Wong-Zakai-type approximation would be needed, and so we obtain uniform convergence in probability in that case. $\square$

The previous theorem provided a general result on convergence in probability which relies solely on topological closeness of strategies. Differently in spirit, an approximation idea due to [BB04] shows that one can actually approximate the proceeds of any strategy almost surely by some cleverly constructed continuous f.v. strategies which can be implemented within the same time interval, if the base price $\overline{S}$ is continuous.

Proposition 2.2.12 (Almost sure uniform approximation *à la* Bank-Baum by continuous f.v. strategies)**.** *Suppose that $\overline{S}$ is continuous and $g(x,\cdot)$ and h are continuously differentiable with locally Hölder-continuous derivatives for some index $\delta > 0$. For any predictable càdlàg process Θ on $[0,T]$ and any $\varepsilon > 0$, there exists a continuous process Θ^ε with f.v. paths such $Y_T^\Theta = Y_T^{\Theta^\varepsilon}$, $\Theta_0^\varepsilon = \Theta_{0-}$ and $|L_T(\Theta) - L_T(\Theta^\varepsilon)| \vee |\Theta_T - \Theta_T^\varepsilon| \leq \varepsilon$, $\mathbb{P}$-a.s.*

Proof. Note that $K(y,t) := G(\overline{S},y) - h(y)\int_0^t g(\overline{S}_u,y)\,\mathrm{d}\langle M\rangle_u$ and $\widetilde{K}(y,t) := h(y)\langle M\rangle_t$ define smooth families of semimartingales in the sense of [BB04, Def. 2.2] and

$$L_T(\Theta) = \int_0^T K(Y_{s-},\mathrm{d}s) - \big(G(\overline{S}_T, Y_T) - G(\overline{S}_{0-}, Y_{0-})\big). \tag{2.22}$$

Predictability of Θ implies predictability of Y and hence Y_T is $\mathcal{F}_{T-}$ measurable. By a slight extension of the approximation result in [BB04, Thm. 4.4] for the non-linear integrator $(K,\widetilde{K})$ (see Lemma 2.5.5 for more details), for every $\varepsilon > 0$ there exists a predictable process Y^ε with continuous paths of finite variation, such that $Y_0^\varepsilon = Y_{0-}$, $Y_T^\varepsilon = Y_T$ and $\mathbb{P}$-a.s.

$$\sup_{0\leq t\leq T}\Big\{\Big|\int_0^t K(Y_{s-},\mathrm{d}s) - \int_0^t K(Y^\varepsilon_{s-},\mathrm{d}s)\Big| \vee \Big|\int_0^t h(Y_s)\,\mathrm{d}\langle M\rangle_s - \int_0^t h(Y_s^\varepsilon)\,\mathrm{d}\langle M\rangle_s\Big|\Big\} \leq \varepsilon.$$

The process Y^ε corresponds to a predictable process Θ^ε with continuous f.v. paths, namely $\Theta^\varepsilon = Y^\varepsilon - Y_{0-} + \Theta_{0-} + \int_0^\cdot h(Y^\varepsilon)\,\mathrm{d}\langle M\rangle_u$, that satisfies $|\Theta_T - \Theta_T^\varepsilon| \leq \varepsilon$, and with reference to (2.22), also satisfies $|L_T(\Theta) - L_T(\Theta^\varepsilon)| \leq \varepsilon$. $\square$

2.2.3 Connection to the Marcus canonical equation

Here we explain briefly, how our proceeds functional connects with an interesting SDE which is known as the Marcus canonical equation [Mar81]. Stability in the sense of Wong–Zakai approximations for this kind of equations has been studied in [KPP95]. Their techniques offer an alternative way to derive the approximation result of Corollary 2.2.10. Recently, stability of such equations for a p-variation rough paths variant of the M_1 topology has been studied in [FC18].

Definition 2.2.13 (Marcus canonical equation)**.** *Let $\Phi : \mathbb{R}^d \to \mathbb{R}^{d\times k}$ be continuously differentiable and Z be a k-dimensional semimartingale. Then the notation*

$$X_t = X_{0-} + \int_0^t \Phi(X_s) \circ \mathrm{d}Z_s \tag{2.23}$$

means that X satisfies the stochastic integral equation

$$\begin{aligned} X_t =& X_{0-} + \int_0^t \Phi(X_{s-})\,\mathrm{d}Z_s + \frac{1}{2}\sum_{j,m=1}^{k}\sum_{\ell=1}^{d}\int_0^t \frac{\partial \Phi_{\cdot,j}}{\partial x_\ell}(X_{s-})\Phi_{\ell,m}(X_{s-})\,\mathrm{d}[Z^j, Z^m]^c_s \\ &+ \sum_{0\le s\le t\,,\,\Delta Z_s\neq 0} \big(\varphi(\Phi(\cdot)\Delta Z_s, X_{s-}) - X_{s-} - \Phi(X_{s-})\Delta Z_s\big), \end{aligned} \tag{2.24}$$

where $\Phi_{\cdot,j}$ is the j^{th} column of Φ, Z^j is the j^{th} entry of Z and $\varphi(\xi, x)$ denotes the value $y(1)$ of the solution to

$$y'(u) = \xi(y(u)) \quad \textit{with} \quad y(0) = x. \tag{2.25}$$

The quadratic (co-)variation process is denoted by $[\cdot] = [\cdot]^c + [\cdot]^d$, it decomposes into a continuous part (appearing in (2.24)) and a discontinuous part. The next lemma gives a representation of the impact and proceeds processes of our model in terms of a Marcus canonical equation for the case $h \in C^1$. To this end, let the function $\Phi : \mathbb{R}^3 \to \mathbb{R}^{3\times 3}$ for $X = (X^1, X^2, X^3)^{tr} \in \mathbb{R}^3$ be given by

$$\Phi(X) := \begin{pmatrix} -g(X^3, X^2) & 0 & 0 \\ 1 & 0 & -h(X^2) \\ 0 & 1 & 0 \end{pmatrix}. \tag{2.26}$$

Lemma 2.2.14. *Let Θ be a càdlàg process with paths of finite total variation, and L be defined by (2.6) be the process describing the evolution of proceeds generated by Θ. Set $X_t := \big(L_t, Y_t, \overline{S}_t\big)^{tr}$, so $X_{0-} = \big(0, Y_{0-}, \overline{S}_{0-}\big)^{tr}$, and $Z_t := \big(\Theta_t, \overline{S}_t, \langle M\rangle_t\big)^{tr}$. Then the process X is the solution to the Marcus canonical equation*

$$X_t = X_{0-} + \int_0^t \Phi(X_s) \circ \mathrm{d}Z_s\,.$$

For the proof see Section 2.5. Following [KPP95, Sect. 6], we now derive a Wong-Zakai-type approximation result in our setup. For a bounded semimartingale process Θ and $\varepsilon > 0$ consider the approximating absolutely continuous processes defined by

$$\Theta_t^\varepsilon := \frac{1}{\varepsilon}\int_{t-\varepsilon}^t \Theta_s \,\mathrm{d}s, \quad t \geq 0, \tag{2.27}$$

with the convention that $\Theta_t = \Theta_{0-}$ for $t < 0$. See Figure 2.1, where $\varepsilon = 1/n$.

Let $Z_t^\varepsilon := (\Theta_t^\varepsilon, \overline{S}_t, \langle M\rangle_t)^{tr}$ and X^ε be a solution to the following SDE in the Itô sense

$$\mathrm{d}X_t^\varepsilon = \Phi(X_t^\varepsilon)\,\mathrm{d}Z_t^\varepsilon, \qquad X_0^\varepsilon = X_{0-}. \tag{2.28}$$

The next result on Wong-Zakai-type convergence is based on the theory from [KPP95, Sect. 5].

Theorem 2.2.15. *Suppose that $\overline{S}$ is continuous and let $(\Theta_t)_{t\geq 0}$ be a bounded semimartingale. For $\varepsilon > 0$, let Θ^ε be the Wong-Zakai-type approximations from (2.27). Let X^ε be defined by (2.28) for $Z_t^\varepsilon := (\Theta_t^\varepsilon, \overline{S}_t, \langle M\rangle_t)^{tr}$ and Φ as in (2.26). For time-changes $\gamma_\varepsilon(t) := \frac{1}{\varepsilon}\int_{t-\varepsilon}^t ([\Theta]_s^d + s)\,\mathrm{d}s$, consider the processes $(\mathcal{X}_t^\varepsilon)_{t\geq 0}$ defined by $\mathcal{X}_t^\varepsilon := X^\varepsilon_{\gamma_\varepsilon^{-1}(t)}$. For $\varepsilon \to 0$ the processes $\mathcal{X}^\varepsilon$ then converge in probability in the compact uniform topology to a process $(\mathcal{X}_t^0)_{t\geq 0}$, such that $X_t = (X_t^1, X_t^2, X_t^3)^{tr} := \mathcal{X}^0_{\gamma_0(t)}$ is a solution of*

$$X_t = X_{0-} + \int_0^t \Phi(X_s) \circ \mathrm{d}Z_s - \Big(\frac{1}{2}\int_0^t g_x(\overline{S}_s, X_{s-}^2)\,\mathrm{d}[\overline{S},\Theta]_s,\ 0,\ 0\Big)^{tr}, \tag{2.29}$$

where $X_{0-} = (0, Y_{0-}, \overline{S}_0)^{tr}$ and $\gamma_0(t) := [\Theta]_t^d + t$.

The proof of Theorem 2.2.15 is delegated to Section 2.5. Theorem 2.2.15 directly gives, since $X^1 = L$, that for a bounded semimartingale strategy Θ, the proceeds $L = L(\Theta)$ of this strategy up to $T < \infty$ take the form

$$\begin{aligned} L_T = &-\int_0^T g(\overline{S}_{t-}, Y_{t-}^\Theta)\,\mathrm{d}\Theta_t - \frac{1}{2}\int_0^T g_y(\overline{S}_{t-}, Y_{t-}^\Theta)\,\mathrm{d}[\Theta]_t^c - \int_0^T g_x(\overline{S}_t, Y_{t-}^\Theta)\,\mathrm{d}[\overline{S},\Theta]_t \\ &- \sum_{\substack{\Delta\Theta_t\neq 0\\ t\leq T}} \Big(\int_0^{\Delta\Theta_t} g(\overline{S}_t, Y_{t-}^\Theta + x)\,\mathrm{d}x - g(\overline{S}_t, Y_{t-}^\Theta)\Delta\Theta_t\Big), \end{aligned} \tag{2.30}$$

where the stochastic integral is understood in Itô's sense and Y^Θ is given as in (2.2). It is straightforward to see that (2.30) coincides with (2.15).

Remark 2.2.16. a) Note that boundedness of Θ implies that X^2 is bounded. Localizing along $\overline{S}$ (the variable X^3), we can assume that g is globally Lipschitz continuous. This implies absolute convergence of the infinite sum in (2.24), see [KPP95, p. 356]. In particular, (2.30) is well-defined.

b) The additional covariation term in the limiting equation (2.29) arises since only the strategies Θ are approximated in a Wong–Zakai sense, but not also unaffected price $\overline{S}$ and clock $\langle M\rangle$. For strategies Θ being of finite variation (as it would be natural under proportional transaction costs), this additional covariation term clearly vanishes.

c) Note that Theorem 2.2.15 implies the results in Corollary 2.2.10 for bounded semimartingale processes Θ. Indeed, Theorem 2.2.15 gives for the first components $L^\varepsilon = X^{\varepsilon,1}$, $L = X^{0,1}$ that for any $\eta > 0$ and any horizon $T \in [0,\infty)$ we have $\mathbb{P}\big[\sup_{t\le T}\big|L^\varepsilon_{\gamma_\varepsilon^{-1}(\gamma_0(t))} - L_t\big| \le \eta\big] \to 1$ for $\varepsilon \to 0$. Since $\gamma_\varepsilon^{-1}(\gamma_0(t)) \to t$ at continuity points of γ_0 (which are the continuity points of Θ and thus of L) it follows that $\mathbb{P}[\Omega^\eta_\varepsilon] \to 1$ as $\varepsilon \to 0$ with

$$\Omega^\eta_\varepsilon := \{\omega \mid \forall t \text{ with } \Delta L_t(\omega) = 0 : |L^\varepsilon_t(\omega) - L_t(\omega)| \le \eta\}.$$

d) The proof of Theorem 2.2.15 could be adapted to the case when M is quasi-left continuous if the bounded semimartingale Θ is assumed to be predictable.

2.3 Absence of arbitrage for the large trader

On the one hand the large trader is faced with adverse price reaction to her trades. On the other hand, her market influence might give her opportunities to manipulate price dynamics in her favor. It is therefore relevant to show that the model does not permit arbitrage opportunities for the large trader in a (fairly large) set of trading strategies. For this section we consider a multiplicative price impact model where $g(\overline{S}, Y) = f(Y)\overline{S}$ with a non-negative, increasing and continuously differentiable function f, cf. Example 2.1.1.[3] Consider a portfolio (β_t, Θ_t) of the large investor, where β_t represents holdings in the bank account (riskless numéraire with discounted value 1) and Θ_t denotes holdings in the risky asset S at time t. We will consider bounded càdlàg strategies Θ on the full time horizon $[0,\infty)$ although our results below will deal with a finite but arbitrary horizon. For the strategy (β, Θ) to be self-financing, the bank account evolves according to

$$\beta_t = \beta_{0-} + L_t(\Theta)\,, \quad t \ge 0, \tag{2.31}$$

with $L(\Theta)$ as in (2.15). In order to define the wealth dynamics induced by the large trader's strategy, we have to specify the dynamics of the value of the risky asset position in the portfolio. If the large trader were to unwind her risky asset position at time t immediately by selling Θ_t shares (meaning to buy shares in case of a short position $\Theta_t < 0$), the resulting change in the bank account would be given by a term of the form (2.5). In this sense, let the *instantaneous liquidation value* process of her position be

$$V^\Theta_t = \beta_t + \overline{S}_t \int_0^{\Theta_t} f(Y^\Theta_t - x)\,\mathrm{d}x\,, \quad t \ge 0. \tag{2.32}$$

This corresponds to the asymptotically realizable real wealth process in [BB04]. Its dynamics (2.33) are mathematically tractable and relevant, e.g. to study no-arbitrage. For $F(x) := \int_0^x f(y)\,\mathrm{d}y$ we have $\overline{S}_t \int_0^{\Theta_t} f(Y^\Theta_t - x)\,\mathrm{d}x = \overline{S}_t\big(F(Y^\Theta_t) - F(Y^\Theta_t - \Theta_t)\big)$. By (2.15) and (2.31), noting that $Y^\Theta - \Theta$ and $\langle M\rangle$ are absolutely continuous processes, we

[3] For additive dynamics of $\overline{S}$ instead of (2.1), one could carry out the analysis in this section also in the case of additive impact $g(\overline{S}, Y) = \overline{S} + f(Y)$

have

$$\begin{aligned}
\mathrm{d}V_t^\Theta &= F(Y_{t-}^\Theta)\,\mathrm{d}\overline{S}_t - \overline{S}_t(fh)(Y_{t-}^\Theta)\,\mathrm{d}\langle M\rangle_t - \mathrm{d}\big(\overline{S}.F(Y_\cdot^\Theta - \Theta.)\big)_t \\
&= \big(F(Y_{t-}^\Theta) - F(Y_{t-}^\Theta - \Theta_{t-})\big)\,\mathrm{d}\overline{S}_t - \overline{S}_t\big(F'(Y_{t-}^\Theta) - F'(Y_{t-}^\Theta - \Theta_{t-})\big)h(Y_{t-}^\Theta)\,\mathrm{d}\langle M\rangle_t \\
&= \big(F(Y_{t-}^\Theta) - F(Y_{t-}^\Theta - \Theta_{t-})\big)\overline{S}_{t-}(\mu_t\,\mathrm{d}\langle M\rangle_t + \mathrm{d}M_t), \qquad (2.33)
\end{aligned}$$

with $\mu_t := \xi_t - h(Y_{t-}^\Theta)\cdot\dfrac{F'(Y_{t-}^\Theta) - F'(Y_{t-}^\Theta - \Theta_{t-})}{F(Y_{t-}^\Theta) - F(Y_{t-}^\Theta - \Theta_{t-})}\mathbb{1}_{\{\Theta_{t-}\neq 0\}}$ and $V_0^\Theta = \beta_0 + \int_0^{\Theta_0} f(Y_0 + x)\,\mathrm{d}x$.

We will prove a no-arbitrage theorem for the large trader essentially for models that do not permit arbitrage opportunities for small investors in the absence of trading by the large trader. More precisely, for this section we assume for the driving noise M the

Assumption 2.3.1. *For every predictable and bounded process μ and every $T \geq 0$, there exists a probability measure $\mathbb{P}^\mu \approx \mathbb{P}$ on $\mathcal{F}_T$ such that the process $M + \int_0^\cdot \mu_s\,\mathrm{d}\langle M\rangle_s$ is a $\mathbb{P}^\mu$-local martingale on $[0,T]$.*

Example 2.3.2 (Models satisfying assumption Assumption 2.3.1)**.** a) If M is continuous, then under our model assumptions from Section 2.1, for every predictable and bounded process μ the probability measure $\mathrm{d}\mathbb{P}^\mu = \mathcal{E}(-\int_0^\cdot \mu_s\,\mathrm{d}M_s)\,\mathrm{d}\mathbb{P}$ is well-defined (thanks to Novikov's condition) and satisfies Assumption 2.3.1.

b) Let M be a Lévy process that is a martingale with $\Delta M > -1$ and $\mathbb{E}[M_1^2] < \infty$. In this case, it is a special semimartingale with characteristic triplet $(0, \sigma, K)$ (w.r.t. the identity truncation function), and we have the decomposition $M = \sqrt{\sigma}W + x * (\mu^M - \nu^{\mathbb{P}})$, where W is a $\mathbb{P}$-Brownian motion (or null if $\sigma = 0$), μ^M is the jump measure of M and $\nu^{\mathbb{P}}(\mathrm{d}x, \mathrm{d}t) = K(\mathrm{d}x)\,\mathrm{d}t$ is the $\mathbb{P}$-predictable compensator of μ^M. We have $\langle M\rangle_t = \lambda t$, $t \geq 0$, for some $\lambda \geq 0$. In the case $\sigma > 0$, Assumption 2.3.1 is clearly satisfied. Indeed, an equivalent change of measure by the standard Girsanov's theorem with respect to the non-vanishing (scaled) Brownian motion M^c can be done such that $M^c + \int \mu\,\mathrm{d}\langle M\rangle$ becomes a martingale, without changing the Lévy measure.

Otherwise, in case of $\sigma = 0$, M is a pure jump Lévy process. For this case, let us restrict our consideration to the situation of two-sided jumps, since pure-jump Lévy processes of such type appear more relevant to the modeling of financial returns than those ones with one-sided jumps only; examples are the exponential transform of the variance-gamma process or the so-called CGMY-process (suitably compensated to give a martingale exponential transform), cf. [KS02, CGMY02] for the relevant notions and models respectively. Here, it turns out that $K((-\infty, 0)) > 0$ and $K((0, +\infty)) > 0$ is already a sufficient condition for Assumption 2.3.1 to hold, i.e. possibility for jumps occurring in both directions. Indeed, a suitable change of measure can then be constructed as follows. Let $n > 0$ be such that $K([1/n, n]) > 0$ and $K([-n, -1/n]) > 0$. Denote $C^+ := \int_{[1/n,n]} x^2 K(\mathrm{d}x) > 0$ and $C^- := \int_{[-n,-1/n]} x^2 K(\mathrm{d}x) > 0$. Define functions $Y^\pm : \mathbb{R} \to \mathbb{R}$ by $Y^+ := 1$ on $[1/n, n]^c$, $Y^+(x) - 1 := x/C^+$ on $[1/n, n]$, and by $Y^- := 1$ on $[-n, -1/n]^c$, $Y^-(x) - 1 := -x/C^-$ on $[-n, -1/n]$, respectively. Thus

$\int_{\mathbb{R}} x(Y^{\pm}(x)-1)K(\mathrm{d}x) = \pm 1$ and hence , with $\eta := \lambda\mu$, the bounded previsible process

$$Y(\omega,t,x) := \eta_t^-(\omega)(Y^+(x)-1) + \eta_t^+(\omega)(Y^-(x)-1) + 1$$

satisfies $\int_{\mathbb{R}} x(Y(x)-1)K(\mathrm{d}x) = -\eta$. The stochastic exponential $Z := \mathcal{E}((Y-1)*(\mu^L-\nu^{\mathbb{P}}))$ is a strictly positive $\mathbb{P}$-martingale, cf. [ES05, Prop. 5]. So for $T \geq 0$ there is a measure $d\mathbb{P}^\mu = Z_T d\mathbb{P}$ with density process $(Z_t)_{t\leq T}$. By Girsanov's theorem [JS03, Thm. III.3.11], $M - 1/Z_- \cdot \langle M, Z\rangle = M + \int_0^\cdot \mu_u \,\mathrm{d}\langle M\rangle_u$ is a $\mathbb{P}^\mu$-local martingale on $[0,T]$.

The set of *admissible trading strategies* that we consider is

$$\mathcal{A} := \big\{(\Theta_t)_{t\geq 0} \mid \text{bounded, predictable, càdlàg, with } V^\Theta \text{ bounded from below,}\\ \Theta_{0-} = 0, \text{ and such that } \Theta_t = 0 \text{ for } t \in [T,\infty) \text{ for some } T < \infty\big\}.$$

Note that for such a strategy Θ it clearly holds $V^\Theta = \beta$ on $[T,\infty)$, i.e. beyond some bounded horizon $T < \infty$ the liquidation value coincides with the cash holdings β_T. Boundedness from below for V^Θ has a clear economical meaning, while the boundedness of Θ may be viewed as a more technical requirement. It ensures under Assumption 2.3.1 the existence of a strategy-dependent measure $\mathbb{Q}^\Theta \approx \mathbb{P}$ (on $\mathcal{F}_T$) so that V^Θ is a $\mathbb{Q}^\Theta$-local martingale on $[0,T]$. This relies on (2.33) and is at the key idea for the proof for

Theorem 2.3.3. *Under Assumption 2.3.1, the model is free of arbitrage up to any finite time horizon $T \in [0,\infty)$, in the sense that there exists no $\Theta \in \mathcal{A}$ with $\Theta_t = 0$ on $t \in [T,\infty)$ such that for the corresponding self-financing strategy (β,Θ) with $\beta_{0-} = 0$ we have*

$$\mathbb{P}[V_T^\Theta \geq 0] = 1 \quad \textit{and} \quad \mathbb{P}[V_T^\Theta > 0] > 0\,. \tag{2.34}$$

Proof. Recall the SDE (2.33) which describes the liquidation value process V, and note that $V_0 = 0$. For each $\Theta \in \mathcal{A}$ we have that (Θ, Y^Θ) is bounded. Thus, the drift μ is bounded as well because, in the case of $\Theta_{t-} \neq 0$, by the mean value theorem we have

$$\frac{F'(Y_{t-}^\Theta) - F'(Y_{t-}^\Theta - \Theta_{t-})}{F(Y_{t-}^\Theta) - F(Y_{t-}^\Theta - \Theta_{t-})} = \frac{f'(z_1)}{f(z_2)} \quad \text{for some } z_{1,2} \text{ between } Y_{t-}^\Theta \text{ and } Y_{t-}^\Theta - \Theta_{t-},$$

and this is bounded from above because f, f' are continuous and $f > 0$ (so it is bounded away from zero on any compact set). Hence, Assumption 2.3.1 guarantees the existence of $\mathbb{P}^\mu \approx \mathbb{P}$ on $\mathcal{F}_T$ such that V^Θ is a $\mathbb{P}^\mu$-local martingale on $[0,T]$, and since it is also bounded from below, it is a $\mathbb{P}^\mu$-supermartingale, so $E^\mu[V_T^\Theta] \leq V_0^\Theta = 0$. This rules out arbitrage opportunities, as described in (2.34), under any probability $\mathbb{P}$ equivalent to $\mathbb{P}^\mu$ on $\mathcal{F}_T$, for any $T \in [0,\infty)$. □

Remark 2.3.4 (Extension to bid-ask spread)**.** Absence of arbitrage in the model with zero bid-ask spread naturally implies no arbitrage for model extensions with spread, at least when the admissible trading strategies have paths of finite variation. To make this precise, let us model different impact processes Y^{Θ^-} and Y^{Θ^+} from selling and buying, respectively, according to (2.2), and best bid and ask price processes $(S^b, S^a) := \big(f(Y^{\Theta^-})\overline{S}^b, f(Y^{\Theta^+})\overline{S}^a\big)$ with $S^b \le S^a$ for non-increasing Θ^- and non-decreasing Θ^+. Then, the proceeds from implementing (Θ^-, Θ^+) on $[0,T]$ would be

$$-\int_0^T S_t^b \,\mathrm{d}\Theta_t^{-,c} - \int_0^T S_t^a \,\mathrm{d}\Theta_t^{+,c} - \sum_{\substack{0\le t\le T\\ \Delta\Theta_t^-<0}} \overline{S}_t^b \int_0^{\Delta\Theta_t^-} f(Y_{t-}^{\Theta^-}+x)\,\mathrm{d}x - \sum_{\substack{0\le t\le T\\ \Delta\Theta_t^+>0}} \overline{S}_t^a \int_0^{\Delta\Theta_t^+} f(Y_{t-}^{\Theta^+}+x)\,\mathrm{d}x.$$

Now for $\Theta := \Theta^- + \Theta^+$, the initial relation $Y_{0-}^{\Theta^-} \le Y_{0-}^{\Theta} \le Y_{0-}^{\Theta^+}$ implies $Y^{\Theta^-} \le Y^{\Theta} \le Y^{\Theta^+}$. Hence $S^b \le S \le S^a$ and the proceeds above for the model with non-vanishing spread would be dominated (a.s.) by those that we get in (2.6), i.e. in the model without bid-ask spread. In an alternative but different variant, one could extend the zero bid-ask spread model to a one-tick-spread model, motivated by insights in [CdL13], by letting $(S^b, S^a) := (S, S+\delta)$ for some $\delta > 0$. Again, proceeds in this model would be dominated by those in the zero-spread model. In either variant, absence of arbitrage opportunities in the zero bid-ask spread model implies the same for an extended model with spread.

Remark 2.3.5 (Extension to càglàd strategies)**.** For any càglàd (left continuous with right limits) $(\Theta_t)_{t\ge0}$ (with $\Theta_{0-} = \Theta_0$) the unique càglàd solution Y^Θ to the integral equation $Y_t - Y_s = \int_s^t h(Y_u)\alpha_u\,du + \Theta_t - \Theta_s$ ($0 \le s < t$, with $Y_0 = Y_{0-}$), corresponding to (2.2), can be defined pathwise (cf. proof of [PTW07, Thm. 4.1]); statements on càdlàg paths ($\bar\Theta$,$Y^{\bar\Theta}$) translate to càglàd paths (Θ,Y^Θ) by relations $\bar\Theta_{t-} = \Theta_t$ and $Y_{t-}^{\bar\Theta} = Y_t^\Theta$, $t \ge 0$. Using this, we can define the dynamics of the liquidation wealth process V for any strategy Θ which is adapted with càglàd paths or predictable with càdlàg paths, and hence locally bounded, by the the unique (strong) solution to the SDE (2.33) for given initial condition $V_0 \in \mathbb{R}$. Thereby, the result on absence of arbitrage can be extended to a larger set of strategies, which contains the set $\mathcal{A}$ and in addition all bounded adapted and càglàd (left-continuous with right limits) processes $(\Theta_t)_{t\ge0}$ with $\Theta_{0-} = \Theta_0 = 0$ for which there exists some $T < \infty$ such that $\Theta_t = 0$ for $t \in [T,\infty)$ holds. Indeed, the same lines of proof show that such Θ cannot give an arbitrage opportunity in the sense of Theorem 2.3.3.

2.4 Application examples and extensions

In this section, we present application examples and model extensions, mostly in the framework of multiplicative impact $g(\overline{S}, Y) = f(Y)\overline{S}$, cf. Example 2.1.1, that highlight different questions in which our stability results are helpful and show the flexibility of our analysis and its applicability to other model. First we briefly discuss in Section 2.4.1 the connection of the asymptotically realizable proceeds from block trades to LOBs. Section 2.4.2 shows, by compactness argument, the existence of an optimal control by an application of our continuity result in Theorem 2.2.7. For this, it is rather easy to check that the set of controls is compact for the M_1 topology. In Section 2.4.3 we identify the solution of an optimal liquidation problem with the already known optimizer in a smaller class of admissible controls, by approximating semimartingale strategies with strategies of bounded variation, where stability of the proceeds functional plays a crucial role.

Sections 2.4.4 and 2.4.5 illustrate modifications of the price impact model by changing the impact process to allow partially instantaneous impact, respectively incorporate permanent impact, to which the analysis in Section 2.2 carries over. Herein, the M_1 topology is again key for identifying the (asymptotically realizable) proceeds and thus extending the models to a larger class of trading strategies. Section 2.4.6 gives an extension of our setup that allows for stochasticity in the impact and will be the setup of Chapter 3.

2.4.1 Limit order book perspective for multiplicative market impact

Multiplicative price impact and the proceeds from block trading can be interpreted by trading in a shadow limit order book (LOB). We now show how the multiplicative price impact function f is related to a LOB shape that is specified in terms of *relative* price pertubations $\rho_t := S_t/\overline{S}_t$, whereas additive impact corresponds to a LOB shape being specified with respect to absolute price pertubations $S_t - \overline{S}_t$ as in [PSS11]. Note that the LOB shape is static. Such can be viewed as a low-frequency model for price impact according to a LOB shape which is representative on longer horizons, but not for high frequency trading over short periods.

Let $s = \rho\overline{S}_t$ be some price close to the unaffected price $\overline{S}_t$ and let $q(\rho)\,\mathrm{d}\rho$ denote the density of (bid or ask) offers at price level s, i.e. at the relative price perturbation ρ. This leads to a measure with cumulative distribution function $Q(\rho) := \int_1^\rho q(x)\,\mathrm{d}x$, $\rho \in (0,\infty)$. The total volume of orders at prices corresponding to perturbations ρ from some range $R \subset (0,\infty)$ then is $\int_R q(x)\,\mathrm{d}x$. Selling $-\Delta\Theta_t$ shares at time t shifts the price from $\rho_{t-}\overline{S}_t$ to $\rho_t\overline{S}_t$, while the volume change is $Q(\rho_{t-}) - Q(\rho_t) = -\Delta\Theta_t$. The proceeds from this sale are $\overline{S}_t \int_{\rho_t}^{\rho_{t-}} \rho\,\mathrm{d}Q(\rho)$. Changing variables, with $Y_t := Q(\rho_t)$ and $f := Q^{-1}$, the proceeds can be expressed as in equation (2.5). In this sense, Y from (2.2) can be understood as the *volume effect process* as in [PSS11, Section 2]. See Figure 2.2 for illustration.

2.4.2 Optimal liquidation problem on finite time horizon

In this example, using continuity of the proceeds in the M_1 topology we will show that the optimal liquidation problem over monotone strategies on a finite time horizon admits an optimal strategy. For $\theta \geq 0$ shares to be liquidated, the problem is to

$$\text{maximize} \quad \mathbb{E}[L_T(\Theta)] \quad \text{over } \Theta \in \mathcal{A}_{\mathrm{mon}}(\theta), \tag{2.35}$$

over the set of all decreasing adapted càdlàg Θ with $\Theta_{0-} = \theta$ and $\Theta\mathbb{1}_{[T,\infty)} = 0$. We consider the situation when the unaffected price process has constant drift, i.e. $\overline{S}_t = e^{\mu t}M_t$ for $t \geq 0$, where $\mu \in \mathbb{R}$ and M is a non-negative continuous martingale that is locally square integrable. Existence and (explicit) structural description of the optimal strategy is already known in the following two cases: a) $\mu = 0$ and any time horizon $T \geq 0$,

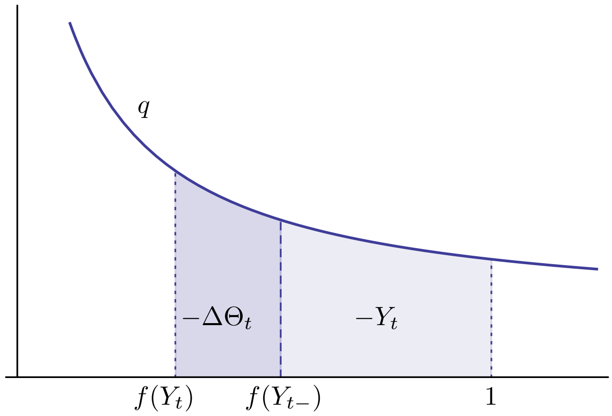


Figure 2.2: Order book density q and behavior of the multiplicative price impact $f(Y)$ when selling a block of size $-\Delta\Theta_t > 0$. Note that $-Y_t = -Y_{t-} - \Delta\Theta_t$.

cf. [PSS11, Løk12]; or: b) $\mu < 0$ and sufficiently big time horizon $T \geq T(\theta, \mu)$ under additional assumptions on f and h, cf. [BBF17a]. There M can be taken even quasi-left continuous in which case the set of admissible strategies should be restricted to predictable processes.

In the general case, the following compactness argument proves existence of an optimizer - without providing any structural description for it, of course. First, it suffices to optimize over deterministic strategies and thus to take $M \equiv 1$ by a change of measure argument, see [BBF17a, Remark 3.9]. Now, for some fixed $\varepsilon > 0$ consider the optimization problem over the set of strategies

$$\widetilde{\mathcal{A}}_{\mathrm{mon}}(\theta) = \{\widetilde{\Theta} \in D[-\varepsilon, T+\varepsilon] \mid \widetilde{\Theta} \text{ is the } \textit{extended path} \text{ of some determ. } \Theta \in \mathcal{A}_{\mathrm{mon}}(\theta)\}.$$

Endowing $\widetilde{\mathcal{A}}_{\mathrm{mon}}(\theta)$ with the Skorokhod M_1 topology makes it relatively compact, which is straightforward to check using [Whi02, Thm. 12.12.2]; the compactness criterion in [Whi02, Thm. 12.12.2] is trivial for such monotone strategies because the M_1 oscillation function is zero and all the paths are constant in neighborhoods of the end points. Thus, if $(\widetilde{\Theta^n}) \subset \widetilde{\mathcal{A}}_{\mathrm{mon}}(\theta)$ is a maximizing sequence (of extended paths) for the problem (2.35), then it (or some subsequence) converges to $\widetilde{\Theta^*} \in D[-\varepsilon, T+\varepsilon]$. By continuity of the proceeds functional L in the M_1 topology (Theorem 2.2.7) we obtain

$$\sup_{\Theta \in \mathcal{A}_{\mathrm{mon}}(\theta)} L_T(\Theta) = \lim_{n\to\infty} L_{T+\varepsilon}(\widetilde{\Theta^n}) = L_{T+\varepsilon}(\widetilde{\Theta^*}). \tag{2.36}$$

Since on $[-\varepsilon, 0)$ (resp. $(T, \varepsilon]$) each $\widetilde{\Theta^n}$ is constant θ (resp. 0) and convergence in M_1 implies local uniform convergence at continuity points of the limit, cf. [Whi02, Lemma 12.5.1], there exists $\Theta^* \in \mathcal{A}_{\mathrm{mon}}(\theta)$ such that $\widetilde{\Theta^*}$ is its extended path in $D[-\varepsilon, T+\varepsilon]$. Thus $L_{T+\varepsilon}(\widetilde{\Theta^*}) = L_T(\Theta^*)$ and Θ^* is an optimal liquidation strategy by (2.36).

2.4.3 Optimal liquidation problem with general strategies

Consider the problem from [BBF17a, Sect. 5] to liquidate a risky asset optimally, posed over the set of bounded variation strategies $\mathcal{A}_{\mathrm{bv}}(\theta)$ with no shortselling, for some initial position $\theta \geq 0$, i.e. $\max_{\Theta \in \mathcal{A}_{\mathrm{bv}}(\theta)} \mathbb{E}[L_\infty(\Theta)]$; Recall that in the setup there the fundamental price process is $\overline{S}_t = e^{-\delta t} M_t$ for some $\delta > 0$ and a non-negative locally square integrable quasi-left continuous martingale M, and $\mathrm{d}\langle M \rangle_t$ in the dynamics of Y in (2.2) is replaced by $\mathrm{d}t$. By [BBF17a, Thm. 5.1], the optimal bounded variation strategy Θ^* is deterministic and liquidates in some finite time $T-1$ (which depends on the model parameters).

Now consider the optimal liquidation problem over the larger set of admissible strategies

$$\mathcal{A}_{\mathrm{semi}}(\theta) := \{\Theta \mid \text{bounded predictable semimartingale, } \Theta \geq 0, \Theta_{0-} = \theta, \Theta_t = \Theta_{t \wedge (T-1)}\}.$$

Note that for any admissible strategy $\Theta \in \mathcal{A}_{\mathrm{semi}}(\theta)$, the (martingale part of the) stochastic integral in equation (2.15) is a true martingale and will vanish in expectation, yielding

$$\mathbb{E}[L_T(\Theta)] = \mathbb{E}\bigg[-\int_0^T e^{-\delta t} M_t((fh)(Y_t^\Theta) + \delta F(Y_t^\Theta))\,\mathrm{d}t - (e^{-\delta T} M_T F(Y_T^\Theta) - M_{0-} F(Y_{0-}^\Theta))\bigg],$$

where $F(x) = \int_0^x f(y)\,\mathrm{d}y$. A change of measure argument as in [BBF17a, Rem. 3.9] shows that we can take w.l.o.g. $M \equiv 1$ and thus it suffices to optimize the proceeds over the set $\mathcal{A}_{\text{càdlàg}}(\theta)$ of all deterministic non-negative càdlàg paths having square-summable jumps, starting at time $0-$ at θ and being zero after time $T-1$. For each such $\Theta \in \mathcal{A}_{\text{càdlàg}}(\theta)$ and every $\varepsilon > 0$, we can find a deterministic bounded variation strategy $\Theta^\varepsilon \in \mathcal{A}_{\mathrm{bv}}(\theta)$ that executes until time T and gives proceeds that are at most ε-away from the proceeds of Θ. Indeed, this follows from Corollary 2.2.10 where the approximating sequence is indeed of bounded variation continuous processes (since Θ is bounded), and noting that the probabilistic nature of the stability results in Section 2.2.2 is due to the presence of the (intrinsically probabilistic) stochastic integral in (2.15), cf. the proof of Theorem 2.2.7, which would be immaterial here in the case of constant M. In particular,

$$\sup_{\Theta \in \mathcal{A}_{\mathrm{semi}}(\theta)} \mathbb{E}[L_T(\Theta)] \leq \sup_{\mathcal{A}_{\text{càdlàg}}(\theta)} \mathbb{E}[L_T(\Theta)] = \sup_{\Theta \in \mathcal{A}_{\mathrm{bv}}(\theta)} \mathbb{E}[L_T(\Theta)] = \mathbb{E}[L_T(\Theta^*)],$$

meaning that Θ^* is optimal also within in the (larger) set $\mathcal{A}_{\mathrm{semi}}(\theta)$.

2.4.4 Price impact with partially instantaneous recovery

This example is inspired by work of [Roc11] on a different (additive impact, block-shaped limit order book (LOB)) price impact model; adapting his interesting idea to our setup leads to an extension of our transient impact model, where a further parameter $\eta \in (0, 1]$ permits for partially instantaneous recovery of price impact. Further, the example illustrates how proceeds from trading could, at first, be given for simple strategies only, and continuity arguments are key for an extension to a larger space of strategies.

Motivated by observations that other traders respond quickly to market orders by adding limit orders in opposite direction, [Roc11] has proposed a model where impact from a block trade is partially instantaneous and partially transient. A market sell (resp. buy) order eats into the bid (resp. ask) side of a LOB and is filled at respective prices, price impact being a function of the shape of the LOB. A certain fraction $1-\eta$ ($0<\eta\le 1$) of that impact is instantaneously recovered directly after the trade, while only the remaining η-fraction constitutes a transient impact that decays gradually over time (cf. (2.37)). As stated in [Roc11], this means that "we think of $1-\eta$ as the fraction of the order book which is renewed after a market order so that in practice the actual impact on prices is η times the full impact". In our previous model for a two-sided LOB (non-monotone strategies), with the idealizing assumption of zero bid-ask spread, the model with full impact ($\eta=1$) implicitly postulates that the gap between bid and ask prices after a block buy (resp. sell) order is filled up instantaneously with ask (resp. bid) orders. For one-directional trading such hypothesis is conservative, but for trading in alternating directions it may be overly optimistic. So, it appears to be an interesting generalization to postulate that the gap is closed from both sides in a certain fraction.

To incorporate this into our setup, let $\eta\in[0,1]$ and suppose that the impact directly after completion of a block trade of size $\Delta\Theta_t$ at time $t\in[0,\infty)$ is actually $Y_{t-}+\eta\Delta\Theta_t$, where Y_{t-} is the market impact immediately before the trade. Thus, the market impact process $Y^{\eta,\Theta}$ evolves according to

$$\mathrm{d}Y^{\eta,\Theta}_t = -h(Y^{\eta,\Theta}_t)\,\mathrm{d}\langle M\rangle_t + \eta\,\mathrm{d}\Theta_t, \quad t\ge 0. \tag{2.37}$$

Indeed, (2.37) holds for simple strategies Θ and hence for all càdlàg trading programs Θ by continuity of $\Theta\mapsto Y^{\eta,\Theta}$ in the uniform and Skorokhod J_1 and M_1 topologies.

The case $\eta=0$ corresponds to no (non-instantaneous) impact while $\eta=1$ gives our previous setup with full impact. The situation where $\eta\in(0,1)$ is more delicate, in that executing a block order at once would always be suboptimal, whereas subdividing a block trade into smaller ones and executing them one after the other would lead to smaller expenses, i.e. larger proceeds, due to the instantaneous partial recovery of price impact. Thus, there would be a difference between *asymptotically realizable* proceeds from a block trade (in the terminology of [BB04]) and its direct proceeds from a LOB interpretation.

Motivated by optimization questions like the optimal trade execution problem where a trader tries to evade illiquidity costs from large (block) orders, if possible, our aim is to specify a model that is stable with respect to small intertemporal changes, in particular approximating block trades by subdividing the trade into small packages and executing them in short time intervals. Thus, the proceeds that we will derive here will be asymptotically realizable. First, let us only assume that at every time $t\ge 0$, the average price per share for a block trade of size Δ is some value between $f(Y_{t-})\overline{S}_t$ and $f(Y_{t-}+\Delta)\overline{S}_t$, where Y_{t-} is the state of the impact process right before the block trade. Hence, the arguments in the proof of Lemma 2.2.1 carry over (with $c=1/\eta$, $Y=Y^\eta/\eta$ and suitably re-scaled functions f,h) and yield that the proceeds from implementing a continuous finite variation strategy Θ should be given by $\tilde{L}_T(\Theta)=-\int_0^T \overline{S}_t f(Y^{\eta,\Theta}_t)\,\mathrm{d}\Theta_t$,

$T \geq 0$, irrespective of a particular initial specification for proceeds from block trades. As such was the starting point for Section 2.2, the analysis there for the case $\eta = 1$ carries over to the model extension for $\eta \in (0,1]$: For any continuous f.v. process Θ we obtain

$$\tilde{L}_T(\Theta) = \frac{1}{\eta}\left(\int_0^T F(Y_{u-}^{\eta,\Theta})\,\mathrm{d}\overline{S}_u - \int_0^T \overline{S}_u (fh)(Y_u^{\eta,\Theta})\,\mathrm{d}\langle M\rangle_u - \big(\overline{S}_T F(Y_T^{\eta,\Theta}) - \overline{S}_0 F(Y_{0-}^{\eta,\Theta})\big)\right). \tag{2.38}$$

By Theorem 2.2.7 the right-hand side of (2.38) is continuous in the predictable strategy Θ taking values in $D([0,T];\mathbb{R})$ when endowed with any of the uniform, Skorokhod J_1 and M_1 topologies. So, asymptotically realizable proceeds are given by (2.38). In particular, asymptotically realizable proceeds from a block sale of size $\Delta \neq 0$ at time t are

$$-\frac{1}{\eta}\overline{S}_t\big(F(y_{t-} + \eta\Delta) - F(y_{t-})\big) = -\frac{1}{\eta}\overline{S}_t \int_0^{\eta\Delta} f(y_{t-} + x)\,\mathrm{d}x\,,$$

where y_{t-} denotes the state of the market impact process before the trade. Note that these proceeds strictly dominate the proceeds $-\overline{S}_t \int_0^{\Delta} f(y_{t-} + x)\,\mathrm{d}x$ that would arise from a executing the block sale in the LOB corresponding to the price impact function f. Also this model variant is free of arbitrage in the sense of Theorem 2.3.3, whose proof carries over. In mathematical terms one may observe, maybe surprisingly, that the model structure (see (2.37) and (2.38)) for the extension $\eta \in (0,1]$ is like the one for the previous model (with $\eta = 1$), and is hence amenable to a likewise analysis. In finance terms, to model partially instantaneous recovery in such a way thus has quantitative effects. But it does not lead to new qualitative features for the model, since the large investor could side-step much of the, at first sight, highly disadvantageous effect from large block trades by trading continuously (in approximation), at least in absence of further frictions.

2.4.5 Incorporating persistent permanent impact

So far the impact in our modelling setup was completely transient, i.e. prices will eventually recover towards the fundamental prices. However, a part of the impact might be persistent, i.e. trading actions could affect the full future dynamics of prices in a way that would not wear off due to resilience. This permanent component of the impact is typically a function of the holdings Θ in the risky asset. To incorporate this in our setup, the following extension is quite natural: the risky asset price is

$$S = g(\overline{S}, Y^{\Theta}, \Theta),$$

where $g : \mathbb{R}^3 \to \mathbb{R}$ is a suitable price impact function. To demonstrate the flexibility of our analysis so far, we will now derive the asymptotically realizable proceeds for such specification of impact that has both transient and permanent component.

Indeed, consider the function $G:\mathbb{R}^3\to\mathbb{R}$ defined by

$$G(\bar{s},y,\theta) := \int_0^y g(\bar{s},u,\theta-y+u)\,\mathrm{d}u, \quad \bar{s},y,\theta\in\mathbb{R}. \tag{2.39}$$

The function G satisfies $G_y + G_\theta = g$. Thus, provided that $G\in C^{2,1,1}$, we get for for continuous f.v. Θ the following equivalent representation for the proceeds functional $L(\Theta) = -\int_0^\cdot g(\overline{S}_u, Y_u^\Theta, \Theta_u)\,\mathrm{d}\Theta_u$:

$$\begin{aligned} L(\Theta) = &\int_0^\cdot G_{\bar{s}}(\overline{S}_{u-}, Y_{u-}^\Theta, \Theta_{u-})\,\mathrm{d}\overline{S}_u \\ &+ \int_0^\cdot \Big(\tfrac{1}{2}G_{\bar{s}\bar{s}}(\overline{S}_u, Y_u^\Theta, \Theta_u)\overline{S}_u^2 - G_y(\overline{S}_u, Y_u^\Theta, \Theta_u)h(Y_u^\Theta)\Big)\,\mathrm{d}\langle M\rangle_u \\ &- \big(G(\overline{S}_\cdot, Y_\cdot^\Theta, \Theta_\cdot) - G(\overline{S}_0, Y_{0-}^\Theta, \Theta_{0-})\big) \\ &+ \sum_{\substack{\Delta\overline{S}_u\neq 0\\ 0\le u\le\cdot}} \big(G(\overline{S}_u, Y_u^\Theta, \Theta_u) - G(\overline{S}_{u-}, Y_u^\Theta, \Theta_u) - G_{\bar{s}}(\overline{S}_{u-}, Y_u^\Theta, \Theta_u)\Delta\overline{S}_u\big)\,, \end{aligned} \tag{2.40}$$

Now, provided that g is non-negative and $(y,\theta)\mapsto G_{\bar{s}\bar{s}}(\bar{s},y,\theta)$ is Lipschitz continuous on compacts, it is straightforward to see that the proof of Theorem 2.2.7 extends to the proceeds functional in (2.40). Thus, L from (2.40) gives the asymptotically realizable proceeds in this case of both permanent and transient impact.

The construction of a function G so that the integration by parts argument to get to the representation (2.40) works was crucial for our analysis. In this chapter, we considered one risky asset and thus an explicit construction of G like in (2.39) is always possible. However, this is not the case in multi-dimensional setup of multi-asset models. As it will turn out from our analysis in Chapter 5, in the multi-asset case the existence of such G, for which a suitable form of the proceeds like in (2.40) could be found, will be equivalent to the absence of profitable asymptotically instantaneous round trips, that are quick round trips that yield positive proceeds.

2.4.6 Market impact with stochastic liquidity

In Chapter 3, we will consider the following extension of our setup that incorporates stochasticity in the volume imbalances modelled by the market impact process Y. Our unaffected price process there follows the dynamics in (2.1) with $M=\sigma W$, where $\sigma>0$ and W is a standard Brownian motion, $\xi_t=\mu\in\mathbb{R}$ is a constant. The impact process Y^Θ is defined by

$$\mathrm{d}Y_t^\Theta = -\beta Y_t^\Theta\,\mathrm{d}t + \hat{\sigma}\,\mathrm{d}B_t + \mathrm{d}\Theta_t, \quad Y_{0-}^\Theta = y, \tag{2.41}$$

where $\beta>0$ and B is a standard Brownian motion correlated with W, i.e. for some $\rho\in[-1,1]$ we have $[W,B]_t=\rho t$ for $t\ge 0$. This additional noise in the dynamics of Y^Θ renders the liquidity stochastic in that the volume imbalances from trading recover in time but randomly.

We will be interested in the discounted proceeds from monotone strategies in the multiplicative impact specification, that is $g(\overline{S}, Y) = \overline{S} f(Y)$ for a suitable function f. In this case, for continuous finite variation Θ the γ-discounted proceeds will be

$$L(\Theta) = -\int_0^\cdot e^{-\gamma t} \overline{S}_t f(Y_t^\Theta) \, \mathrm{d}\Theta_t. \tag{2.42}$$

Integration by parts gives the following equivalent form of L :

$$\begin{aligned} L(\Theta) = &\int_0^\cdot F(Y_t^\Theta) \, \mathrm{d}(e^{-\gamma t} \overline{S}_t) - \int_0^\cdot e^{-\gamma t} \overline{S}_t f(Y_t^\Theta)(h(Y_t^\Theta) - \sigma \hat{\sigma} \hat{\rho}) \, \mathrm{d}t \\ &+ \int_0^\cdot \hat{\sigma} e^{-\gamma t} \overline{S}_t f(Y_t^\Theta) \, \mathrm{d}B_t - e^{-\gamma \cdot} \overline{S} F(Y^\Theta) \big|_{0-}^{\cdot}, \end{aligned} \tag{2.43}$$

where $F(x) = \int_0^x f(u) \, \mathrm{d}u$.

In this case the stability result in Theorem 2.2.7 holds true also for the current specifications of Y^Θ from (2.41) and L from (2.43) respectively. Indeed, its proof could be easily adapted to the current setup since we still have continuity of $\Theta \mapsto Y^\Theta$ and moreover L is structurally unchanged. Indeed, continuity of $\Theta \mapsto Y^\Theta$ follows from Proposition 2.5.1 because the map $\Theta \mapsto \Theta + \hat{\sigma} B$ is continuous in all considered topologies by continuity of the paths of $\hat{\sigma} B$ (needed for the J_1 and M_1 topology), and Y^Θ in (2.41) is "driven" by $\Theta \mapsto \Theta + \hat{\sigma} B$. Therefore, L from (2.43) gives the asymptotically realizable proceeds as the continuous extension of L from (2.42) to general adapted càdlàg strategies (cf. also Remark 2.2.8). In particular, if Θ is of finite variation, L has the following equivalent representation that will be the basis for our analysis in Chapter 3: for $T \geq 0$

$$L_T(\Theta) = -\int_0^T e^{-\gamma t} \overline{S}_t f(Y_t^\Theta) \, \mathrm{d}\Theta_t^c - \sum_{\substack{\Delta\Theta_t \neq 0 \\ 0 \leq t \leq T}} e^{-\gamma t} \overline{S}_t \int_0^{\Delta\Theta_t} f(Y_{t-} + x) \, \mathrm{d}x,$$

where $\Theta = \Theta^c + \sum_{\substack{\Delta\Theta_t \neq 0 \\ 0 \leq t \leq \cdot}} \Delta\Theta_t$ is the pathwise decomposition of Θ into a continuous and a pure jump part.

2.5 Some auxiliary proofs

The next proposition collects known continuity properties of the solution map $\Theta \mapsto Y^\Theta$ on $D([0,T];\mathbb{R})$ from (2.2), with the presentation being adapted to our setup.

Proposition 2.5.1. *Assume that h is Lipschitz continuous and $\langle M \rangle = \int_0^\cdot \alpha_s \, \mathrm{d}s$ with pathwise (locally) Lipschitz density α. Then the solution map $D([0,T];\mathbb{R}) \to D([0,T];\mathbb{R})$, with $\Theta \mapsto Y^\Theta$ from* (2.2)*, is defined pathwise. The map is continuous when the space $D([0,T];\mathbb{R})$ is endowed with either the uniform topology or the Skorokhod J_1 or M_1 topology. Moreover, if Θ is an adapted càdlàg process, then the process Y^Θ is also* *adapted.*

Proof. The proof in the case of the uniform topology and the Skorokhod J_1 topology is given in [PTW07, proof of Thm. 4.1]; the proof there is for $\alpha \equiv 1$ but it clearly extends to our setup as long as α is Lipschitz. For the M_1 topology, cf. [PW10, Thm. 1.1], where again the main argument ([PW10, proof of Thm. 1.1]) extends to our setup of more general α. That Y^Θ is adapted follows from the (pathwise) construction of Y^Θ as the (a.s.) limit (in the uniform topology) of adapted processes, the solution processes for a sequence of piecewise-constant controls Θ^n approximating uniformly Θ, cf. [PTW07, proof of Thm. 4.1]. □

In general, we may have $\alpha_n \to \alpha$ and $\beta_n \to \beta$ in $D([0,T])$ endowed with J_1 (or M_1), and yet $\alpha_n + \beta_n \not\to \alpha + \beta$ when α and β have a common jump time. However, in special cases like in what follows, this does not happen.

Lemma 2.5.2 (Allowed cancellation of jumps for J_1)**.** *Let $\alpha_n \to \alpha_0$ and $\beta_n \to \beta_0$ in $(D([0,T]), J_1)$ with the following property: for every $n \geq 0$ and every $t \in (0,T)$*

$$\Delta\alpha_n(t) \neq 0 \quad \textit{implies} \quad \Delta\beta_n(t) = -\Delta\alpha_n(t).$$

Then $\alpha_n + \beta_n \to \alpha_0 + \beta_0$ in $(D([0,T]), J_1)$.

Proof. By [JS03, Prop. VI.2.2, a] it suffices to check that for every $t \in (0,T)$ there exists a sequence $t_n \to t$ such that $\Delta\alpha_n(t_n) \to \Delta\alpha_0(t)$ and $\Delta\beta_n(t_n) \to \Delta\beta_0(t)$.

Let $t \in (0,T)$ be arbitrary and first suppose that $\Delta\alpha_0(t) \neq 0$. Then [JS03, Prop. VI.2.1, a] implies the existence of a sequence $t_n \to t$ such that $\Delta\alpha_n(t_n) \to \Delta\alpha_0(t)$. Thus, our assumption on the sequence (β_n) gives $\Delta\beta_n(t_n) \to \Delta\beta_0(t)$. For the case $\Delta\alpha_0(t) = 0$, let $t_n \to t$ be such that $\Delta\beta_n(t_n) \to \Delta\beta_0(t)$. By [JS03, Prop. VI.2.1, b.5] we conclude that $\Delta\alpha_n(t_n) \to \Delta\alpha_0(t)$ as well, finishing the proof. □

Let us note that the conclusion of Lemma 2.5.2 does not hold for the M_1 topology. Consider for example $\alpha_0 = \mathbb{1}_{[1,\infty)}$ with approximating sequence $\alpha_n(t) := n\int_t^{t+1/n} \alpha_0(s)\,\mathrm{d}s$ and $\beta_0 = 1 - \alpha_0$ with approximating sequence $\beta_n(t) := n\int_{t-1/n}^{t} \beta_0(s)\,\mathrm{d}s$. Thus we need the following refined statement.

Lemma 2.5.3 (Allowed cancellation of jumps for M_1)**.** *Let $\alpha_n \to \alpha_0$ in $(D([0,T]), \|\cdot\|_\infty)$ and $\beta_n \to \beta_0$ in $(D([0,T]), M_1)$ with the following property: $t \in Disc(\alpha_0)$ implies $\beta_n \to \beta_0$ locally uniformly in a neighborhood of t. Then $\alpha_n + \beta_n \to \alpha_0 + \beta_0$ in $(D([0,T]), M_1)$.*

Proof. We prove the following claim that suffices to deduce M_1-convergence of $\alpha_n + \beta_n$: For any $t \in [0,T]$ and $\varepsilon > 0$ there are $\delta > 0$ and $n_0 \in \mathbb{N}$ such that

$$w_s(\alpha_n + \beta_n, t, \delta) \leq w_s(\alpha_n, t, \delta) + w_s(\beta_n, t, \delta) + \varepsilon \quad \text{for all } n \geq n_0, \tag{2.44}$$

where w_s is the M_1 oscillation function, see [Whi02, Chap. 12, eq. (4.4)]. Indeed, if (2.44) holds, then the second condition in [Whi02, Thm. 12.5.1(v)] would hold, while the first condition there holds because of local uniform convergence at points of continuity of $\alpha_0 + \beta_0$: Either there is cancellation of jumps and thus local uniform convergence by our assumption, or both paths do not jump which still gives local uniform convergence because M_1-convergence implies such at continuity points of the limit.

To check (2.44), we have $\lim_{\delta\downarrow 0}\limsup_{n\to\infty} v(\alpha_n,\alpha_0,t,\delta)=0$ at points $t\in[0,T]$ with $\Delta\alpha_0(t)=0$, where for $x_1,x_2\in D([0,T])$

$$v(x_1,x_2,t,\delta) := \sup_{0\vee(t-\delta)\le t_1,t_2\le(t+\delta)\wedge T} |x_1(t_1)-x_2(t_2)|,$$

see [Whi02, Thm. 12.4.1], which implies (2.44) for small δ and large n. Now if $t\in \mathrm{Disc}(\alpha_0)$, $\alpha_n\to\alpha_0$ and $\beta_n\to\beta_0$ locally uniformly in a neighborhood of t which implies that for small δ and large n

$$w_s(\alpha_n+\beta_n,t,\delta)\le w_s(\alpha_0+\beta_0,t,\delta)+\varepsilon/2.$$

Because $\alpha_0+\beta_0\in D([0,T])$, we can make $w_s(\alpha_0+\beta_0,t,\delta)$ smaller than $\varepsilon/2$, which finishes the proof. $\square$

Lemma 2.5.4 (Uniform convergence of jump term). *Let $\alpha,\beta_n,\beta\in D([0,T])$ be such that $[\alpha]^d_T := \sum_{t\le T:\Delta\alpha(t)\neq 0}|\Delta\alpha(t)|^2<\infty$, β_n are uniformly bounded and at every jump time $t\in[0,T]$ of α, $\Delta\alpha(t)\neq 0$, we have pointwise convergence $\beta_n(t)\to\beta(t)$. Let $G\in C^2$ such that $y\mapsto G_{xx}(x,y)$ is Lipschitz continuous on compacts. Then the sum*

$$J(\alpha,\beta_n)_t := \sum_{\substack{u\le t\\ \Delta\alpha(t)\neq 0}} G\big(\alpha(t),\beta_n(t)\big)-G\big(\alpha(t-),\beta_n(t)\big)-G_x\big(\alpha(t-),\beta_n(t)\big)\Delta\alpha(t)$$

converges uniformly for $t\in[0,T]$ to $J(\alpha,\beta)_t$, as $n\to\infty$.

Proof. Since α, $[\alpha]^d$, β_n and β are bounded on $[0,T]$ by a constant $C\in\mathbb{R}$, we can assume w.l.o.g. that G_{xx} is globally Lipschitz in y with Lipschitz constant L. Hence $J(\alpha,\beta_n)_t<\infty$ by Taylor's theorem. Let $H(x,\Delta x,y):=G(x+\Delta x,y)-G(x,y)-G_x(x,y)\Delta x$ and denote by $\tilde J^{n,\pm}$ the increasing and decreasing components of $J(\alpha,\beta_n)-J(\alpha,\beta)$, respectively, i.e.

$$\tilde J^{n,+}_t := \sum_{\substack{u\le t\\ \tilde H(\dots)>0}} \tilde H\big(\alpha(u-),\Delta\alpha(u),\beta_n(u),\beta(u)\big),\ \tilde J^{n,-}_t := \sum_{\substack{u\le t\\ \tilde H(\dots)<0}} \tilde H\big(\alpha(u-),\Delta\alpha(u),\beta_n(u),\beta(u)\big),$$

for $\tilde H(x,\Delta x,y,z):=H(x,\Delta x,y)-H(x,\Delta x,z)$. Moreover, take any enumeration of the jump times of α, $\{t_k\mid k\in\mathbb{N}\}=\{t\mid\Delta\alpha(t)\neq 0\}$, and arbitrary $\varepsilon>0$. Since $[\alpha]^d<\infty$, there exists $K\in\mathbb{N}$ such that $\sum_{k>K}|\Delta\alpha(t_k)|^2<\varepsilon/(2CL)$. Moreover, we have $|\tilde H(x,\Delta x,y,z)|\le\frac12|\Delta x|^2L|y-z|$ and thus

$$|\tilde J^{n,\pm}_T|\le\frac{L}{2}\sum_{k=1}^{\infty}|\Delta\alpha(t_k)|^2|\beta_n(t_k)-\beta(t_k)|<\frac{\varepsilon}{2}+\frac{L}{2}\Big(\max_{1\le k\le K}|\beta_n(t_k)-\beta(t_k)|\Big)\sum_{k=1}^{K}|\Delta\alpha(t_k)|^2\,.$$

By pointwise convergence $\beta_n(t_k)\to\beta(t_k)$ at all t_k, there exists $N\in\mathbb{N}$ such that for all

$k = 1, \dots, K$ and $n \geq N$ we have $|\beta_n(t_k) - \beta(t_k)| < \varepsilon/(L[\alpha]^d_T)$ and therefore $|\tilde{J}^{n,\pm}_T| < \varepsilon$ for $n \geq N$. Hence $J^{n,\pm}_T \to 0$ as $n \to \infty$.

Since $J^{n,\pm}$ are monotone and do not cross zero, we have $\sup_{0 \leq t \leq T} |\tilde{J}^{n,\pm}_t| = |\tilde{J}^{n,\pm}_T|$ and therefore uniform convergence $\tilde{J}^{n,\pm} \to 0$ on $[0, T]$. So in particular $J(\alpha, \beta_n)$ converges to $J(\alpha, \beta)$, uniformly on $[0, T]$. □

The next statement extends the approximation result [BB04, Thm. 4.4] on non-linear integrators to a smooth family of $\mathbb{R}^2$ valued semimartingales, needed in Proposition 2.2.12.

Lemma 2.5.5. *Let $(L^y) = (L^y_1, L^y_2)$ ($y \in \mathbb{R}$) be an $\mathbb{R}^2$-valued smooth family of continuous semimartingales. Let Y be an L-integrable predictable process and fix $Y_0 \in L^0(\mathcal{F}_0)$, $Y_T \in L^0(\mathcal{F}_{T-})$. Then, for any $\varepsilon > 0$, there exists a predictable process Y^ε with continuous paths of finite variation such that $Y^\varepsilon_0 = Y_0$, $Y^\varepsilon_T = Y_T$, and*

$$\sup_{t \in [0,T]} \left| \int_0^t L_i(Y^\varepsilon_u, \mathrm{d}u) - \int_0^t L_i(Y_u, \mathrm{d}u) \right| \leq \varepsilon \quad \textit{for } i = 1, 2,\ \mathbb{P}\textit{-a.s.}$$

Proof. Since the proof follows closely the arguments in [BB04], we just sketch them with outlining the differences. The Borel-Cantelli argument in the proof of [BB04, Thm. 4.4] could be applied here as well. More precisely, for $\varepsilon_n = \varepsilon/2^n$, $n \geq 0$, set $\tau_0 = 0$, $Y^\varepsilon_0 = Y_0$. The construction of Y^ε is done inductively as follows. Assume that Y^ε is constructed already on the interval $[\![0, \tau_n]\!]$. Take now a strategy $Y^{n+1} = Y^{\varepsilon_{n+1}, \tau_n, Y^\varepsilon_{\tau_n}}$ whose paths are of finite variation over $[\![\tau_n, T]\!]$, such that $Y^{\varepsilon_{n+1}, \tau_n, Y^\varepsilon_{\tau_n}}_{\tau_n} = Y^\varepsilon_{\tau_n}$, $Y^{\varepsilon_{n+1}, \tau_n, Y^\varepsilon_{\tau_n}}_T = Y_T$ and

$$\mathbb{P}\left[\sup_{\tau_n \leq t \leq T} \left| \int_{\tau_n}^t L_i(Y^{\varepsilon_{n+1}, \tau_n, Y^\varepsilon_{\tau_n}}_u, \mathrm{d}u) - \int_{\tau_n}^t L_i(Y_u, \mathrm{d}u) \right| \geq \varepsilon_{n+1} \right] \leq \frac{\varepsilon_{n+1}}{2}, \quad \text{for } i = 1, 2.$$

The existence of such process will be argued below. Choose now the stopping time

$$\tau_{n+1} := \inf \left\{ t \geq \tau_n \ : \ \max_{i=1,2} \left| \int_{\tau_n}^t L_i(Y_u, \mathrm{d}u) - \int_{\tau_n}^t L_i(Y^{n+1}_u, \mathrm{d}u) \right| > \varepsilon_{n+1} \right\} \wedge T$$

and extend continuously the definition of Y^ε from $[\![0, \tau_n]\!]$ to $[\![0, \tau_{n+1}]\!]$ by setting $Y^\varepsilon = Y^{n+1}$ on $]\!]\tau_n, \tau_{n+1}]\!]$. Since by construction of Y^{n+1} we have $\mathbb{P}[\tau_{n+1} < T] \leq \varepsilon_{n+1}$, a Borel-Cantelli argument applies and gives that we can carry out only a finite number of the above inductive steps in order to define on $[0, T]$ the adapted process Y^ε with continuous paths of finite variation. Moreover, Y^ε satisfies for $i = 1, 2$,

$$\sup_{0 \leq t \leq T} \left| \int_0^t L_i(Y_u, \mathrm{d}u) - \int_0^t L_i(Y^\varepsilon_u, \mathrm{d}u) \right| \leq \sum_{n=0}^\infty \varepsilon_n = \varepsilon.$$

Hence it only remains to justify the existence of the processes $Y^{\varepsilon_{n+1}, \tau_n, Y^\varepsilon_{\tau_n}}$ from above. This is done exactly like in the proof of [BB04, Lemma A.1]. There, by continuity of the non-linear integrals in the ucp topology, one first argues that it suffices to consider Y being a simple process. Afterwards, linearly interpolating on a finer grid (of size $\Delta > 0$) one gets a continuous piecewise linear approximation Y^Δ of Y, making use of the fact that any $\mathcal{F}_{T-}$-measurable random variable is the terminal value of some continuous adapted process with piecewise linear paths of bounded variation, see [BB04, Lemma A.2] for details. To conclude, a dominated convergence argument together with the Burkholder-Davis-Gundy inequality give that for this piecewise linear approximation Y^{Δ} of Y, the non-linear integrals $\int_0 L_i(Y^{\Delta}_u,\,\mathrm{d}u)$ converge to $\int_0 L_i(Y_u,\,\mathrm{d}u)$ in the ucp topology, as $\Delta \to 0$. The precise details are given in the proof of [BB04, Lemma A.1]. □

Next we provide the proofs for the results in Section 2.2.3.

Proof of Lemma 2.2.14. Since Θ is of finite variation, we have $\mathrm{d}[Z^j, Z^m]^c_t = \mathrm{d}[\overline{S}]^c_t$ for $j = m = 2$, and 0 otherwise. So the $\partial\Phi_{\cdot,j}/\partial x_\ell$ terms in equation (2.24) simplify to

$$\frac{1}{2}\sum_{j,m=1}^{3}\sum_{\ell=1}^{3}\int_0^t \frac{\partial\Phi_{\cdot,j}}{\partial x_\ell}(X_{s-})\Phi_{\ell,m}(X_{s-})\,\mathrm{d}[Z^j, Z^m]^c_s = (0,\ 0,\ 0)^{tr}. \tag{2.45}$$

Jumps of Z are of the form $\Delta Z_s = \big(\Delta\Theta_s,\ \Delta\overline{S}_s,\ 0\big)^{tr}$, so for $\xi(X) := \Phi(X)\Delta Z_s$ we obtain $\xi(X) = \big(-g(X^3, X^2)\Delta\Theta_s,\ \Delta\Theta_s,\ \Delta\overline{S}_s\big)^{tr}$, which yields the solutionto (2.25) as $y(u) = V_u = (V^1_u, V^2_u, V^3_u)^{tr} \in \mathbb{R}^3$ with $V_0 = X_{s-}$,

$$\begin{aligned}
V^2_u &= Y_{s-} + \int_0^u \Delta\Theta_s\,\mathrm{d}x = Y_{s-} + u\Delta\Theta_s\,,\\
V^3_u &= \overline{S}_{s-} + \int_0^u \Delta\overline{S}_s\,\mathrm{d}x = \overline{S}_{s-} + u\Delta\overline{S}_s\,,\\
V^1_u &= L_{s-} - \int_0^u g(\overline{S}_{s-} + x\Delta\overline{S}_s, Y_{s-} + x\Delta\Theta_s)\Delta\Theta_s\,\mathrm{d}x\\
&= L_{s-} - \int_0^{u\Delta\Theta_s} g(\overline{S}_{s-}, Y_{s-} + x)\,\mathrm{d}x\,,
\end{aligned}$$

since quasi-left continuity of $\overline{S}$ gives that a.s. $\Delta\overline{S}_s = 0$ whenever $\Delta\Theta_s \neq 0$ (jumps of Θ occur at predictable times). Thus the jump terms in (2.24) become

$$\begin{aligned}
&\varphi(\Phi(\cdot)\Delta Z_s, X_{s-}) - X_{s-} - \Phi(X_{s-})\Delta Z_s\\
&\qquad = \left(-\int_0^{\Delta\Theta_s} g(\overline{S}_{s-}, Y_{s-} + x)\,\mathrm{d}x + g(\overline{S}_{s-}, Y_{s-})\Delta\Theta_s,\ 0,\ 0\right)^{tr}.
\end{aligned} \tag{2.46}$$

Furthermore, the Itô integral in (2.24) reads

$$\int_0^t \Phi(X_{s-})\,\mathrm{d}Z_s = \begin{pmatrix} -\int_0^t g(\overline{S}_{s-}, Y_{s-})\,\mathrm{d}\Theta_s \\ -\int_0^t h(Y_s)\,\mathrm{d}\langle M\rangle_s + \Theta_t - \Theta_{0-} \\ \overline{S}_t - \overline{S}_{0-} \end{pmatrix}. \tag{2.47}$$

Summing up X_{0-} and equations (2.45) to (2.47) yields the second and third components $Y_{0-} - \int_0^t h(Y_s)\,\mathrm{d}s + \Theta_t - \Theta_{0-} = Y_t$ and $\overline{S}_{0-} + \overline{S}_t - \overline{S}_{0-} = \overline{S}_t$, respectively. To complete the proof, we note that for the first component we get

$$L_{0-}-\int_0^t g(\overline{S}_{s-},Y_{s-})\,\mathrm{d}\Theta_s+\sum_{\substack{0\le s\le t\\ \Delta\Theta_s\neq 0}}\Big(g(\overline{S}_{s-},Y_{s-})\Delta\Theta_s-\int_0^{\Delta\Theta_s} g(\overline{S}_{s-},Y_{s-}+x)\,\mathrm{d}x\Big)=L_t\,.\ \square$$

Proof of Theorem 2.2.15. The proof follows the ideas in [KPP95, Section 6] where the statement is proved for one-dimensional SDEs, whereas here we need a multidimensional version. For convenience of the reader, we will indicate the changes in the arguments to accommodate our setup. Localizing along $\overline{S}$ (the variable X^3), we can assume that Φ, its partial derivatives and products $(\partial\Phi_{\cdot,j}/\partial x_\ell)\Phi_{\ell,m}$ are globally Lipschitz continuous and bounded. The localized solutions can be easily pasted together because of the global existence and strong uniqueness of a solution to the Marcus SDE $\mathrm{d}X_t=\Phi(X_t)\circ\mathrm{d}Z_t$ in our case; note also that the localizing sequence will not affect the time-changes γ_ε which additionally simplifies the argument. Now let $V_t^\varepsilon := Z^\varepsilon_{\gamma_\varepsilon^{-1}(t)}$. Then $\mathcal{X}^h$ is the unique solution of $\mathcal{X}^\varepsilon = X_{0-}+\int_0^t\Phi(\mathcal{X}_s^\varepsilon)\,\mathrm{d}V_s^\varepsilon$. The analysis is complicated by the fact that the sequence (V^ε) is not *good*, see [KPP95, Section 6] for definition. For this reason we rewrite the equation for $\mathcal{X}^\varepsilon$ in the following form, keeping in mind that $(Z_{\gamma_\varepsilon^{-1}})$ is a good sequence of semimartingales:

$$\mathcal{X}_t^\varepsilon = X_{0-}+\int_0^t\Phi(\mathcal{X}_s^\varepsilon)\,\mathrm{d}Z_{\gamma_\varepsilon^{-1}(s)}+\int_0^t\Phi(\mathcal{X}_s^\varepsilon)\,\mathrm{d}(V_s^\varepsilon - Z_{\gamma_\varepsilon^{-1}(s)}). \tag{2.48}$$

Note that $V_t^\varepsilon - Z_{\gamma_\varepsilon^{-1}(t)} = \big((\Theta^\varepsilon-\Theta)_{\gamma_\varepsilon^{-1}(t)},0,0\big)^{tr} =: (U_t^\varepsilon,0,0)^{tr}$. For the limit, we have $\lim_{\varepsilon\searrow 0}U_t^\varepsilon = \overline{\Theta}_t-\Theta_{\gamma_0^{-1}(t)} =: U_t$, where $\overline{\Theta}_t$ is the limit (in the compact uniform topology) of $\Theta^\varepsilon_{\gamma_\varepsilon^{-1}}$ given by

$$\overline{\Theta}_t := \begin{cases}\Theta_{\gamma_0^{-1}(t)} & \text{if } \eta_1(t)=\eta_2(t),\\ \Theta_{\gamma_0^{-1}(t)}\cdot\frac{t-\eta_1(t)}{\eta_2(t)-\eta_1(t)}+\Theta_{\gamma_0^{-1}(t)-}\cdot\frac{\eta_2(t)-t}{\eta_2(t)-\eta_1(t)} & \text{if } \eta_1(t)\neq\eta_2(t),\end{cases}$$

where $\eta_1(t):=\sup\{s\mid \gamma_0^{-1}(s)<\gamma_0^{-1}(t)\}$ and $\eta_2(t):=\inf\{s\mid\gamma_0^{-1}(s)>\gamma_0^{-1}(t)\}$; as in [KPP95, Lemma 6.2]. The last term in (2.48) is $\big(-\int_0^t g(\overline{S}_{\gamma_\varepsilon^{-1}(s)},\mathcal{X}_s^{\varepsilon,2})\mathrm{d}U_s^\varepsilon, U_t^\varepsilon,0\big)^{tr}$. To identify the limit of the first component, we integrate by parts to obtain

$$\begin{aligned}&\int_0^t g(\overline{S}_{\gamma_\varepsilon^{-1}(s)},\mathcal{X}_s^{\varepsilon,2})\,\mathrm{d}U_s^\varepsilon\\ &\quad= g(\overline{S}_{\gamma_\varepsilon^{-1}(t)},\mathcal{X}_t^{\varepsilon,2})U_t^\varepsilon-\int_0^t U_{s-}^\varepsilon\,\mathrm{d}\big(g(\overline{S}_{\gamma_\varepsilon^{-1}(s)},\mathcal{X}_s^{\varepsilon,2})\big)-[g(\overline{S}_{\gamma_\varepsilon^{-1}(\cdot)},\mathcal{X}^{\varepsilon,2}),U^\varepsilon]_t\\ &\quad= g(\overline{S}_{\gamma_\varepsilon^{-1}(t)},\mathcal{X}_t^{\varepsilon,2})U_t^\varepsilon-\int_0^t U_{s-}^\varepsilon g_x(\overline{S}_{\gamma_\varepsilon^{-1}(s)},\mathcal{X}_s^{\varepsilon,2})\,\mathrm{d}\overline{S}_{\gamma_\varepsilon^{-1}(s)}\end{aligned}$$

$$
\begin{aligned}
&- \frac{1}{2}\int_0^t U^\varepsilon_{s-} g_{xx}(\overline{S}_{\gamma_\varepsilon^{-1}(s)}, \mathcal{X}^{\varepsilon,2}_s)\,\mathrm{d}[\overline{S}_{\gamma_\varepsilon^{-1}}]_s \\
&+ \int_0^t U^\varepsilon_{s-} g_y(\overline{S}_{\gamma_\varepsilon^{-1}(s)}, \mathcal{X}^{\varepsilon,2}_s) h(\mathcal{X}^{\varepsilon,2}_s)\,\mathrm{d}\gamma_\varepsilon^{-1}(s) \\
&- \int_0^t U^\varepsilon_{s-} g_y(\overline{S}_{\gamma_\varepsilon^{-1}(s)}, \mathcal{X}^{\varepsilon,2}_s)\,\mathrm{d}\Theta^\varepsilon_{\gamma_\varepsilon^{-1}(s)} - [g(\overline{S}_{\gamma_\varepsilon^{-1}}, \mathcal{X}^{\varepsilon,2}), U^\varepsilon]_t
\end{aligned}
\tag{2.49}
$$

Note that $\int_0^\cdot U^\varepsilon_{s-}\,\mathrm{d}\overline{S}_{\gamma_\varepsilon^{-1}(s)} \Rightarrow \int_0^\cdot U_{s-}\,\mathrm{d}\overline{S}_{\gamma_0^{-1}(s)} \equiv 0$ and $\int_0^\cdot U^\varepsilon_{s-}\,\mathrm{d}\gamma_\varepsilon^{-1}(s) \Rightarrow \int_0^\cdot U_{s-}\,\mathrm{d}\gamma_0^{-1}(s) \equiv 0$, where "$\Rightarrow$" denotes weak convergence of the processes (in the Skorokhod topology J_1 on the path space). Since these sequences are also *good*, the second and the forth term in (2.49) vanish in the limit. The third term also vanishes in the limit since $([\overline{S}_{\gamma_\varepsilon^{-1}}])$ is a *good* sequence as well.

The quadratic covariation process in the last term of (2.49) can be written as

$$
\begin{aligned}
-\big[g(\overline{S}_{\gamma_\varepsilon^{-1}}, \mathcal{X}^{\varepsilon,2}), U^\varepsilon\big]_t &= -\big[g(\overline{S}_{\gamma_\varepsilon^{-1}}, \mathcal{X}^{\varepsilon,2}), (\Theta^\varepsilon - \Theta)_{\gamma_\varepsilon^{-1}(\cdot)}\big]_t \\
&= -\big[g(\overline{S}_{\gamma_\varepsilon^{-1}(\cdot)}, X^{\varepsilon,2}_{\gamma_\varepsilon^{-1}(\cdot)}), (\Theta^\varepsilon - \Theta)_{\gamma_\varepsilon^{-1}(\cdot)}\big]_t \\
&= -\big[g(\overline{S}, X^{\varepsilon,2}), (\Theta^\varepsilon - \Theta)\big]_{\gamma_\varepsilon^{-1}(t)} \\
&= \big[g(\overline{S}, X^{\varepsilon,2}), \Theta\big]_{\gamma_\varepsilon^{-1}(t)} \\
&= \int_0^{\gamma_\varepsilon^{-1}(t)} g_x(\overline{S}_u, X^{\varepsilon,2}_u)\,\mathrm{d}[\Theta, \overline{S}]_u \\
&= \int_0^t g_x(\overline{S}_{\gamma_\varepsilon^{-1}(s)}, \mathcal{X}^{\varepsilon,2}_s)\,\mathrm{d}[\Theta, \overline{S}]_{\gamma_\varepsilon^{-1}(s)}\,,
\end{aligned}
$$

where the time-changed equalities can be justified by [RY99, Proposition V.1.4].

Note that the process $\mathcal{X}^{\varepsilon,2}$ satisfy the ODE $\mathrm{d}\mathcal{X}^{\varepsilon,2}_t = -h(\mathcal{X}^{\varepsilon,2}_t)\,\mathrm{d}\gamma_\varepsilon^{-1}(t) + \mathrm{d}\Theta_{\gamma_\varepsilon^{-1}(t)} + \mathrm{d}U^\varepsilon_t$. Since γ_ε^{-1} is *good* and $(\gamma_\varepsilon^{-1}, \Theta_{\gamma_\varepsilon^{-1}(\cdot)}, U^\varepsilon) \Rightarrow (\gamma_0^{-1}, \Theta_{\gamma_0^{-1}(\cdot)}, U)$ holds, we have $\mathcal{X}^{\varepsilon,2} \Rightarrow \mathcal{X}^{0,2}$ with

$$
\mathrm{d}\mathcal{X}^{0,2}_t := -h(\mathcal{X}^{0,2}_{t-})\,\mathrm{d}\gamma_0^{-1}(t) + \mathrm{d}\Theta_{\gamma_0^{-1}(t)} + \mathrm{d}U_t.
$$

In particular, for the fifth term in (2.49) note that [KPP95, Lemma 6.3] gives

$$
\int_0^t U^\varepsilon_{s-} g_y(\overline{S}_{\gamma_\varepsilon^{-1}(s)}, \mathcal{X}^{\varepsilon,2}_s)\,\mathrm{d}\Theta^\varepsilon_{\gamma_\varepsilon^{-1}(s)} \quad \Rightarrow \quad \frac{1}{2}\int_0^t g_y(\overline{S}_{\gamma_0^{-1}(s)}, \mathcal{X}^{0,2}_s)\,\mathrm{d}((U_s)^2 - [\Theta]_{\gamma_0^{-1}(s)}).
$$

As in [KPP95, Section 6], we conclude that the right-hand side of (2.49) converges in distribution to

$$g(\overline{S}_{\gamma_0^{-1}(t)}, \mathcal{X}_t^{0,2})U_t - \frac{1}{2}\int_0^t g_y(\overline{S}_{\gamma_0^{-1}(s)}, \mathcal{X}_s^{0,2})\,\mathrm{d}\big((U_s)^2 - [\Theta]^d_{\gamma_0^{-1}(s)}\big)$$
$$+\frac{1}{2}\int_0^t g_y(\overline{S}_{\gamma_0^{-1}(s)}, \mathcal{X}_s^{0,2})\,\mathrm{d}[\Theta]^c_{\gamma_0^{-1}(s)} + \int_0^t g_x(\overline{S}_{\gamma_0^{-1}(s)}, \mathcal{X}_s^{0,2})\,\mathrm{d}[\Theta, \overline{S}]_{\gamma_0^{-1}(s)}. \tag{2.50}$$

Let $\{\tau_i\}$ be the jump times of Θ. Note that $[\Theta]^d$ only changes at times τ_i and $U_t = 0$ if $t \notin [\gamma_0(\tau_i-), \gamma_0(\tau_i))$ for any τ_i. Thus, the first line in (2.50) only changes when $t \in [\gamma_0(\tau_i-), \gamma_0(\tau_i))$ for some τ_i. Now, for $t \in [\gamma_0(\tau_i-), \gamma_0(\tau_i)]$ we have that $U_t = -\Delta\Theta_{\tau_i}(\eta_2(t)-t)/(\eta_2(t)-\eta_1(t))$, and so $-\frac{1}{2}\,\mathrm{d}\big((U_t)^2 - [\Theta]^d_{\gamma_0^{-1}(t)}\big) = |\Delta\Theta_{\tau_i}|^{-2}(\gamma_0(\tau_i)-t)\,\mathrm{d}t$ and $\mathcal{X}_t^{0,2} = \mathcal{X}^{0,2}_{\gamma_0(\tau_i-)} + \Delta\Theta_{\tau_i}(t-\eta_1(t))/(\eta_2(t)-\eta_1(t))$. Thus, using $\eta_2(t)-\eta_1(t) = |\Delta\Theta_{\tau_i}|^2$ and integrating by parts we get that the contribution from (2.50) over the full time interval $[\gamma_0(\tau_i-), \gamma_0(\tau_i)]$ is

$$\int_0^{\Delta\Theta_{\tau_i}} g(\overline{S}_{\tau_i}, \mathcal{X}^{0,2}_{\gamma_0(\tau_i-)} + x)\,\mathrm{d}x - g(\overline{S}_{\tau_i}, Y^{0,2}_{\gamma_0(\tau_i-)})\Delta\Theta_{\tau_i}.$$

Note that this is minus the jump term in the definition of the Marcus integral. So, collecting all the intermediate results so far we conclude like in [KPP95, Section 6] that $\mathcal{X}^\varepsilon$ converges in distribution to a process $\mathcal{X}^0$ such that $X_t = \mathcal{X}^0_{\gamma_0(t)}$. Now, the convergence in the compact uniform topology follows from the argument in the proof of [KPP95, Theorem 6.5]. $\square$

3 Optimal liquidation under stochastic liquidity

In this chapter, which corresponds to the article [BBF18b], we study the optimal liquidation problem in a multiplicative price impact model where liquidity is stochastic in that the volume effect process Y, which determines the inter-temporal resilience of the market, is taken to be stochastic, being driven by additional noise. The chapter is organized as follows. Section 3.2 states the solution for the singular stochastic control problem posed in Section 3.1, and outlines the general course of arguments to come. In Section 3.3, a calculus of variations problem is posed by restricting to strategies given by diffusions reflected at smooth boundaries. The free boundary is thereby constructed in Section 3.4. By solving the HJB variational inequality (3.9), we prove optimality and derive the value function and the optimal control in Section 3.5.

3.1 The model and the optimal control problem

We consider a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ with two correlated Brownian motions W and B with correlation coefficient $\rho \in [-1, 1]$, such that

$$[W, B]_t = \rho t\,, \quad t \geq 0.$$

for the quadratic co-variation of W and B. The filtration $(\mathcal{F}_t)_{t\geq 0}$ is assumed to satisfy the usual conditions of completeness and right continuity, so we can take càdlàg versions for semimartingales. For notions from stochastic analysis we refer to [JS03].

We consider a market with a risky asset, in addition to the riskless numéraire asset whose (discounted) price is constant at 1. The large investor holds $\Theta_t \geq 0$ shares of the risky asset at time t. She may liquidate her initial position of Θ_{0-} shares by trading according to

$$\Theta_t := \Theta_{0-} - A_t\,,$$

where A is a predictable, càdlàg, monotone process, describing the cumulative number of assets sold up to time t. We define the set of admissible strategies as

$$\begin{aligned}\mathcal{A}(\Theta_{0-}) := \{A \mid\ & A \text{ non-decreasing, càdlàg, predictable,} \\ & \text{with } 0 := A_{0-} \leq A_t \leq \Theta_{0-}\}.\end{aligned}$$

The unaffected fundamental price $\overline{S} = (\overline{S}_t)_{t\geq 0}$ of the risky asset evolves according to

$$\mathrm{d}\overline{S}_t = \mu \overline{S}_t \, \mathrm{d}t + \sigma \overline{S}_t \, \mathrm{d}W_t \,, \quad \overline{S}_0 \in (0,\infty), \text{ with } \sigma > 0,\, \mu \in \mathbb{R}, \tag{3.1}$$

as a geometric Brownian motion, in the absence of perturbations by large investor trading. By trading, however, the large investor has market impact on the actual price

$$S_t := f(Y_t)\overline{S}_t \,, \tag{3.2}$$

of the risky asset through some impact process Y, by an increasing positive smooth function $f > 0$ with $f(0) = 1$. The process Y can be interpreted as a volume effect process, representing the transient volume displacement by large trades in a limit order book (LOB) whose shape corresponds to the price impact function f, see Remark 3.1.3. For $\hat{\sigma} > 0$ the effect from perturbations $\hat{\sigma}\,\mathrm{d}B_t - \mathrm{d}A_t$ on the process

$$\mathrm{d}Y_t = -\beta Y_t \, \mathrm{d}t + \hat{\sigma} \, \mathrm{d}B_t - \, \mathrm{d}A_t \,, \quad Y_{0-} = y, \tag{3.3}$$

is transient over time, in that Y is mean reverting towards zero with mean reversion rate $\beta > 0$. Existence and uniqueness of a strong solution to (3.3) are guaranteed for instance by [PTW07, Thm. 4.1]. Sometimes we shall write $Y^{y,A}$ to stress the dependence of Y on its initial state y and the strategy A. The dynamics of Y are of Ornstein-Uhlenbeck type, driven by $\hat{\sigma}\,\mathrm{d}B - \mathrm{d}A$. The mean-reversion property of the OU process has the financial interpretation that in the absence of activity from the large trader, the impact lessens since Y reverts back to the neutral state zero and hence the price recovers to the fundamental price $\overline{S}$, thus modeling the transient component of the impact (in absolute terms).

For $\gamma \geq 0$, the γ-discounted proceeds up to time t from a liquidation strategy A are

$$L_t(y;A) := \int_0^t e^{-\gamma u} f(Y_u)\overline{S}_u \, \mathrm{d}A_u^c + \sum_{\substack{0\leq u\leq t \\ \Delta A_u \neq 0}} e^{-\gamma u}\overline{S}_u \int_0^{\Delta A_u} f(Y_{u-} - x)\, \mathrm{d}x, \tag{3.4}$$

for $t \geq 0$, where $A_t = A_t^c + \sum_{u\leq t} \Delta A_u$ is the (pathwise) decomposition of A into its continuous and pure-jump part, and $Y = Y^{y,A}$ solves (3.3). Jump terms in (3.4) can be justified from a LOB perspective (cf. Remark 3.1.3 below) or by our stability results in the previous chapter, see Section 2.4.6 for details.

As L is an increasing process, the limit $L_\infty := \lim_{t\to\infty} L_t$ exists. The large trader's objective is to maximize expected (discounted) proceeds over an infinite time horizon,

$$\max_{A\in\mathcal{A}(\Theta_{0-})} \mathbb{E}[L_\infty(y;A)] \quad \text{ with } \quad v(y,\theta) := \sup_{A\in\mathcal{A}(\theta)} \mathbb{E}[L_\infty(y;A)], \tag{3.5}$$

where $v(y,\theta)$ denotes the value function for $y \in \mathbb{R}$ and $\theta \in [0,\infty)$.

Remark 3.1.1. The value function v is increasing in y and θ. Indeed, monotonicity in θ follows from $\mathcal{A}(\theta_1) \subset \mathcal{A}(\theta_2)$ for $\theta_1 \leq \theta_2$. For monotonicity in y, note that for $y_1 \leq y_2$ and any strategy $A \in \mathcal{A}(\theta)$ one has $Y_t^{y_1,A} \le Y_t^{y_2,A}$ for all t, implying $L_t(y_1;A) \le L_t(y_2;A)$.

For the rest of the chapter, the function f and scalars $\beta, \mu, \gamma, \sigma, \rho, \hat\sigma$ satisfy

Assumption 3.1.2.

C1. *We have $\delta := \gamma - \mu > 0$, that means the drift coefficient $-\delta\overline{S}$ for the γ-discounted fundamental price $\mathrm{e}^{-\gamma t}\overline{S}_t$ is negative.*

C2. *The impact function $f \in C^3(\mathbb{R})$ satisfies $f, f' > 0$ and $(f'/f)' < (\Phi'/\Phi)'$, where*

$$\Phi(x) := \Phi_\delta(x) := H_{-\delta/\beta}\Big((\sigma\rho\hat\sigma - \beta x)/\big(\sqrt{\beta}\hat\sigma\big)\Big), \tag{3.6}$$

with Hermite function H_ν (cf. [Leb72, Sect. 10.2]) and $\sigma, \hat\sigma, \beta > 0$ and $\rho \in [-1,1]$.

C3. *The impact function f furthermore satisfies $(f'/f)' < (\Phi''/\Phi')'$.*

C4. *The function $\lambda(y) := f'(y)/f(y)$, $y \in \mathbb{R}$, is bounded, i.e. there exists $\lambda_{\max} \in (0,\infty)$ such that $0 < \lambda(y) \le \lambda_{\max}$ for all $y \in \mathbb{R}$.*

C5. *The function $k(y) := \frac{\hat\sigma^2}{2}\frac{f''(y)}{f(y)} - (\beta+\delta) + (\sigma\rho\hat\sigma - \beta y)\frac{f'(y)}{f(y)}$ is strictly decreasing.*

C6. *There exist y_0 and $y_\infty \in \mathbb{R}$ such that $(f'/f)(y_0) = (\Phi'/\Phi)(y_0)$ and $(f'/f)(y_\infty) = = (\Phi''/\Phi')(y_\infty)$ holds.*

Assumption 3.1.2 is satisfied by e.g. $f(y) = \exp(\lambda y)$ with $\lambda \in (0,\infty)$, cf. Lemma 3.4.1 below. See [BBF17a, Section 2.1] for the shape of the related multiplicative LOB. Note that Φ is (up to a constant factor) the unique positive and increasing solution of the ODE $\frac{\hat\sigma^2}{2}\Phi''(y) + (\sigma\rho\hat\sigma - \beta y)\Phi'(y) - \delta\Phi(y) = 0$.

The overall negative drift in Assumption C1 ensures that the optimization problem on an infinite time horizon has a finite value. Assumptions C2 and C3 imply uniqueness of the (boundary) points y_0 and y_∞ from Assumption C6 which are needed in Lemma 3.4.3. While C3, uniqueness of y_∞, is not crucial there, it will be needed in (3.45) for the verification. The bound on λ in Assumption C4 is used to show some growth condition on the value function in Lemma 3.5.5, that is required to apply the martingale optimality principle (Proposition 3.5.1). Assumption C5 is needed for the verification Lemma 3.5.7.

Let us now comment on the model and its financial interpretation.

Remark 3.1.3 (Relating the price impact function to a shadow limit order book density)**.** We explain how the price impact function f can be interpreted in terms of a (static) *multiplicative* limit order book (LOB) and Y can be viewed as a *volume effect process* in spirit of [PSS11], which in our context relates the *relative* price impact to transient imbalances of volume. To this end, let us recall the connection between price impact function f and the (general) density of a LOB. For relative price perturbations $r_t := S_t/\overline{S}_t$, let $q(r)\,\mathrm{d}r$ denote the density of offers available at price $r\overline{S}_t$. We call the (signed) measure induced by q the multiplicative limit order book. Its cumulative distribution function is $Q(r) := \int_1^r q(x)\,\mathrm{d}x$. The total volume of assets available for prices in some (interval) range $\{r\overline{S}_t \mid r \in R\}$ with measurable $R \subset (0,\infty)$ is $\int_R q(x)\,\mathrm{d}x$.

So, a block sale of size $\Delta A_t > 0$ at time t moves the price from $r_{t-}\overline{S}_t$ to $r_t\overline{S}_t$ such that the volume changes according to $Q(r_t) = Q(r_{t-}) - \Delta A_t$, giving (discounted) proceeds $e^{-\gamma t}\overline{S}_t \int_{r_t}^{r_{t-}} x \,\mathrm{d}Q(x)$. In the terminology of [Kyl85], $Q(r_t) - Q(r_{t-})$ reflects the *depth* of the LOB for price changes by a factor of r_t/r_{t-}. A change of variables with $Y_t := Q(r_t)$ and $f := Q^{-1}$ yields the jump term in (3.4). In this sense, Y denotes the effect from the past and present trades on the volume displacement in the LOB. By the drift in (3.3), this effect is persistent over time but not permanent. Its transient nature relates to the liquidity property that [Kyl85] calls resilience. Note that in our model the resilience is stochastic in the sense that the volume effect process Y in (3.3) is, whereas the resilience rate β is constant (differently e.g. to [GH17]).

Remark 3.1.4. Stochasticity may account for variations of transient impact that cannot be entirely explained by the single agent's own trading activity, and thus not solely described by deterministic functional modeling.

(a) Most of the literature on transient impact considers impact that is a deterministic function of the actions of a single large trader. We consider here an application problem for an individual large trader, but we do not want to assume that she is the only large trader in the market, or that she is as an aggregate of all large traders (a possibility mentioned in [Fre98]). The additional stochastic noise term $\hat{\sigma}\,\mathrm{d}B_t$ in (3.3) can be understood as the aggregate influence on the impact by other large 'noise' traders (acting non-strategically). Questions on strategic behavior between multiple traders (like in [SZ17]) are interesting but beyond the present thesis.

(b) Note that the volatility and as well the drift of the (marginal) price process $S = f(Y_t)\overline{S}_t$ from (3.2), at which (additional infinitesimally) small quantities of the risky assets would be traded, are stochastic via the additional stochastic component of Y. Furthermore, we emphasize that the form of relative price impact function $\Delta \mapsto f(Y_{t-} + \Delta)/f(Y_{t-})$ can vary with Y in general. In the sense of Remark 3.1.3, this means the general shape of the corresponding LOB can exhibit stochastic variations from the large trader's perspective.

(c) Recently, [LN17] suggested to model a signal, which predicts the short-term evolution of prices, as an Ornstein-Uhlenbeck process that modulates the drift of the price dynamics. One can interpret stochasticity of Y as such a signal as follows. For $\lambda = f'/f$ being constant, the log-price can be written as $\log S = (\log \overline{S} + \lambda Y^{\mathrm{sig}}) + \lambda Y^{\mathrm{trans},\Theta}$, where Y^{sig} is a mean-reverting signal with $\mathrm{d}Y_t^{\mathrm{sig}} = -\beta Y_t^{\mathrm{sig}}\,\mathrm{d}t + \hat{\sigma}\,\mathrm{d}B_t$ and $Y^{\mathrm{trans},\Theta}$ is the transient impact from trading with $\mathrm{d}Y_t^{\mathrm{trans},\Theta} = -\beta Y_t^{\mathrm{trans},\Theta}\,\mathrm{d}t + \mathrm{d}\Theta_t$. From this perspective, the optimal liquidation strategy will be adaptive to the signal and depend on the correlation between the signal and $\log \overline{S}$, see Theorem 3.2.1 and Remark 3.2.3.

Remark 3.1.5 (Level of interpretation for the model and relation to additive impact)**.** Noting that a bid-ask spread is not modeled explicitly and price impact f (i.e. the LOB shape) is static, we consider the model as being at a mesoscopic level for low-frequency problems, rather than for market microstructure effects in high frequency. At this level and as pointed out in [AKS16, Rmk. 2.2], it is sensible to think of price impact and liquidity costs as being aggregated over various types of orders. The LOB from

Remark 3.1.3 should be interpreted accordingly. Note that in this chapter we deal with monotone strategies and thus only one (bid) side of the LOB is relevant. Considering infinite time horizon can be viewed as approximation for a longer horizon with more analytic tractability. Concerning the question of comparison with additive models for transient impact, positivity of asset prices is desirable from a theoretical point of view, relevant for applications with longer time horizons (as they may occur e.g. for large institutional trades, cf. e.g. [CL95], or for hedging problems with longer maturities), and appears to fit better to common models with multiplicative price evolutions like (3.1). See [BBF17a, Example 5.4] for a more detailed discussion and further references.

3.2 The optimal strategy and how it will be derived

This section states the main theorem which describes the solution to the singular stochastic control problem, and outlines afterwards the general course of arguments for proving it in the subsequent sections. To explain ideas, let us first motivate how the variational inequality (3.9), being the dynamical programming equation for the optimization problem at hand, is readily suggested by an application of the martingale optimality principle. To this end, consider for an admissible strategy A the process

$$G_t(y;A) := L_t(y;A) + e^{-\gamma t}\overline{S}_t V(Y_t, \Theta_t), \tag{3.7}$$

where $G_{0-}(y;A) = \overline{S}_0 V(Y_{0-}, \Theta_{0-})$ and $V \in C^{2,1}(\mathbb{R} \times [0,\infty); [0,\infty))$ is some function. Suppose V can be chosen such that G is a supermartingale. Then one should have

$$\begin{aligned}\overline{S}_0 V(y, \Theta_{0-}) &= \mathbb{E}[G_{0-}(y;A)] \\ &\geq \lim_{T\to\infty} \mathbb{E}[L_T(y;A)] + \lim_{T\to\infty} e^{-\gamma T}\mathbb{E}[\overline{S}_T V(Y_T, \Theta_T)] \\ &= \mathbb{E}[L_\infty(y;A)]\end{aligned}$$

heuristically, provided that the second summand on the right-hand side converges to 0. Hence, for V being such that G is a supermartingale for every admissible strategy A and a martingale for at least one strategy A^*, one can conclude that V is essentially the value function for (3.5) (modulo the factor $\overline{S}_0$). To describe V, one may apply Itô's formula to get

$$\begin{aligned}\mathrm{d}G_t = e^{-\gamma t}\overline{S}_t\Big(&\hat{\sigma} V_y(Y_{t-}, \Theta_{t-})\,\mathrm{d}B_t + \sigma V(Y_{t-}, \Theta_{t-})\,\mathrm{d}W_t \\ &+ \big((\mu-\gamma)V + (\sigma\rho\hat{\sigma} - \beta Y_{t-})V_y + \tfrac{\hat{\sigma}^2}{2}V_{yy}\big)(Y_{t-}, \Theta_{t-})\,\mathrm{d}t \\ &+ \big(f - V_y - V_\theta\big)(Y_{t-}, \Theta_{t-})\,\mathrm{d}A_t^c \\ &+ \int_0^{\Delta A_t} \big(f - V_y - V_\theta\big)(Y_{t-} - x, \Theta_{t-} - x)\,\mathrm{d}x\Big).\end{aligned} \tag{3.8}$$

Define, with $\delta = \gamma - \mu$, a differential operator on $C^{2,0}$ functions φ by

$$\mathcal{L}\varphi(y,\theta) := \frac{\hat\sigma^2}{2}\varphi_{yy}(y,\theta) + (\sigma\rho\hat\sigma - \beta y)\varphi_y(y,\theta) - \delta\varphi(y,\theta).$$

By equation (3.8), solving the Hamilton-Jacobi-Bellman (HJB) variational inequality

$$0 = \max\{f - V_y - V_\theta\ ,\ \mathcal{L}V\} \quad \text{with } V(y,0) = 0,\ y \in \mathbb{R}, \tag{3.9}$$

would suffice for G to be a local (super-)martingale. This suggests the existence of a *sell region* $\mathcal{S}$ (action region) where the $\mathrm{d}A$-integrand $f - V_y - V_\theta$ is zero and it is optimal to trade (i.e. sell), and a *wait region* $\mathcal{W}$ (inaction region) in which the $\mathrm{d}t$-integrand $\mathcal{L}V$ is zero and it is optimal not to trade. Assume that the two regions

$$\begin{aligned}\mathcal{S} &= \{(y,\theta) \in \mathbb{R} \times (0,\infty) \mid \mathrm{y}(\theta) < y\} \quad \text{and}\\ \mathcal{W} &= \{(y,\theta) \in \mathbb{R} \times (0,\infty) \mid y < \mathrm{y}(\theta)\}\end{aligned}$$

are separated by a free boundary $\{(y,\theta) \mid y = \mathrm{y}(\theta)\}$. An optimal strategy, i.e. a strategy for which G is a martingale, would be described as follows: if $(Y_{0-}, \Theta_{0-}) \in \mathcal{S}$, then perform a block sale of size ΔA_0 such that $(Y_0, \Theta_0) = (Y_{0-} - \Delta A_0, \Theta_{0-} - \Delta A_0) \in \partial\mathcal{S}$. Thereafter, if $\Theta_0 > 0$, sell just enough as to keep the process (Y, Θ) within $\overline{\mathcal{W}}$. In this way, the process (Y, Θ) should be described by a diffusion process that is reflected at the boundary $\partial\mathcal{W} \cap \partial\mathcal{S}$ in direction $(-1,-1)$, i.e. there is waiting in the interior and selling at the boundary until all shares are sold, when (Y, Θ) hits $\{(y,0) \mid \mathrm{y}(0) \le y\}$. For such reflected diffusions, exsistence and uniqueness follow from classical results, see Remark 3.3.2, and Theorem 3.3.3 provides important characteristics which are key to the subsequent construction of the optimal control. The solution of the optimal liquidation problem is indeed described by the local time process of a diffusion reflected at a boundary which is explicitly given by an ODE. This main result is stated as Theorem 3.2.1 below.

In the following sections, we will find the value function for our stochastic control problem by constructing a classical solution of the variational inequality (3.9). Provided that the key variational inequalities for the (candidate) solution are satisfied, optimality can be verified by typical martingale arguments, see Proposition 3.5.1. Based on results on reflected diffusions from Theorem 3.3.3, we reformulate in Section 3.3 the optimization problem as a (nonstandard) calculus of variations problem. Its solution, derived in Section 3.4, provides a candidate for the free boundary, separating the regions of action and inaction, together with the value function on that boundary. Moreover, we show a (one-sided) local optimality property of the derived boundary (cf. Theorem 3.4.6). This will be crucial in Section 3.5 (cf. proof of Lemma 3.5.7) to verify (3.9) for the candidate value function, constructed there, in order to finally conclude on p. 74 the proof for

Theorem 3.2.1. *Let Assumption 3.1.2 be satisfied. Then the ordinary differential equation*

$$\mathrm{y}'(\theta) = \left(\frac{((\Phi')^2 - \Phi\Phi'')(f'\Phi' - f\Phi'')/\Phi}{(\Phi\Phi'' - (\Phi')^2)f'' + (\Phi'\Phi'' - \Phi\Phi''')f' + (\Phi'\Phi''' - (\Phi'')^2)f}\right)\!\big(\mathrm{y}(\theta)\big) \tag{3.10}$$

with initial condition $\mathrm{y}(0) = y_0$ *admits a unique solution* $\mathrm{y} : [0, \infty) \to \mathbb{R}$*, that is strictly decreasing and maps* $[0, \infty)$ *bijectively to* $(y_\infty, y_0]$*, for* y_0 *and* y_∞ *from Assumption C6.*

The boundary function y *characterizes the solution of problem* (3.5) *as* $A^* = (\Delta + K)\mathbb{1}_{[\![0,\tau]\!]}$*, where* $\Delta := \Theta_{0-}\mathbb{1}_{\{Y_{0-} \geq y_0 + \Theta_{0-}\}} + \tilde{\Delta}\mathbb{1}_{\{Y_{0-} < y_0 + \Theta_{0-}, \tilde{\Delta} \geq 0\}}$ *with* $\tilde{\Delta} \leq \Theta_{0-}$ *satisfying* $Y_{0-} - \tilde{\Delta} = \mathrm{y}(\Theta_{0-} - \tilde{\Delta})$*, and where* (Y, K) *is the unique continuous adapted process on* $[\![0, \tau]\!]$ *with non-decreasing* K *which solves the* y*-reflected SDE*

$$
\begin{aligned}
Y_t &\leq \mathrm{y}(\Theta_{0-} - \Delta - K_t)\,,\\
\mathrm{d}Y_t &= -\beta Y_t \,\mathrm{d}t + \hat{\sigma}\,\mathrm{d}B_t - \mathrm{d}K_t\,,\\
\mathrm{d}K_t &= \mathbb{1}_{\{Y_t = \mathrm{y}(\Theta_{0-} - \Delta - K_t)\}}\,\mathrm{d}K_t\,,
\end{aligned}
$$

starting in $(Y_{0-} - \Delta, 0)$*, for time to liquidation* $\tau := \inf\{t \geq 0 : K_t = \Theta_{0-} - \Delta\}$*. Moreover,* τ *has finite moments.*

Remark 3.2.2. The optimal control A^* acts as follows: 1) If $Y_{0-} \geq y_0 + \Theta_{0-}$, sell everything immediately at time 0 and stop trading; 2) Otherwise, if (Θ_{0-}, Y_{0-}) is such that $\mathrm{y}(\Theta_{0-}) < Y_{0-} < y_0 + \Theta_{0-}$, perform an initial block trade of size $A^*_0 := \Delta > 0$ so that $Y_0 = Y_{0-} - \Delta$ is on the boundary $Y_0 = \mathrm{y}(\Theta_0)$. Now being in the wait region $\overline{\mathcal{W}}$, sell as much as to keep with the least effort the state process (Y, Θ) in $\overline{\mathcal{W}}$ until all assets are liquidated at time τ (cf. Figure 3.1: waiting e.g. at times $t \in [28, 32]$ since then impact Y_t is less than $\mathrm{y}(\Theta_t)$).

The inverse local time $\tau_\ell := \inf\{t > 0 : K_t > \ell\}$ is simply how long it takes to liquidate ℓ assets (after an initial block sale). For $\tau > 0$ (case 2 in Remark 3.2.2) its Laplace transform is

$$
\mathrm{E}\big[e^{-\alpha\tau_\ell}\big] = \frac{\Phi_\alpha(Y_0)}{\Phi_\alpha(\mathrm{y}(\Theta_0))}\exp\bigg(\int_0^\ell \big(\mathrm{y}'(\Theta_0 - x) + 1\big)\frac{\Phi'_\alpha(\mathrm{y}(\Theta_0 - x))}{\Phi_\alpha(\mathrm{y}(\Theta_0 - x))}\,\mathrm{d}x\bigg) \tag{3.11}
$$

for $\alpha > 0$ and $0 \leq \ell \leq \Theta_0 = \Theta_{0-} - \Delta$, as it will be shown in the proof of Theorem 3.2.1. Using analyticity of Φ_α w.r.t. the parameter α, one easily gets that τ_ℓ has finite moments. Moreover, the Laplace transform (3.11) gives access to the distribution of the time to liquidation τ by efficient numerical inversion, as in e.g. [AW95].

Remark 3.2.3. The optimizer depends on the volatility of the fundamental price. If correlation ρ is not zero, the optimal strategy and the shape of the free boundary do depend on the volatility σ of the fundamental price process. This is a notable difference to many additive impact models, where the optimal liquidation strategy does not depend on the martingale part of the fundamental price process, cf. e.g. [LS13, Sect. 2.2]. To stress the dependence on ρ, we write Φ^ρ for Φ in (3.6), denote by F^ρ the right-hand side of (3.10) and by y_0^ρ the root of $f'/f - (\Phi^\rho)'/\Phi^\rho$. So the solution y^ρ of the ODE $(\mathrm{y}^\rho)'(\theta) = F^\rho\big(\mathrm{y}^\rho(\theta)\big)$ with $\mathrm{y}^\rho(0) = y_0^\rho$ is the optimal boundary function from Theorem 3.2.1. In the special case of constant λ, i.e. $f(y) = e^{\lambda y}$, we have $F^\rho(y) = F^0(y - \sigma\rho\hat{\sigma}/\beta)$ since $\Phi^\rho(y) = \Phi^0(y - \sigma\rho\hat{\sigma}/\beta)$, and thus $\mathrm{y}^\rho(\theta) = \mathrm{y}^0(\theta) + \sigma\rho\hat{\sigma}/\beta$. For general f, investigating y_0 and y_∞ from Assumption C6 still reveals a similar

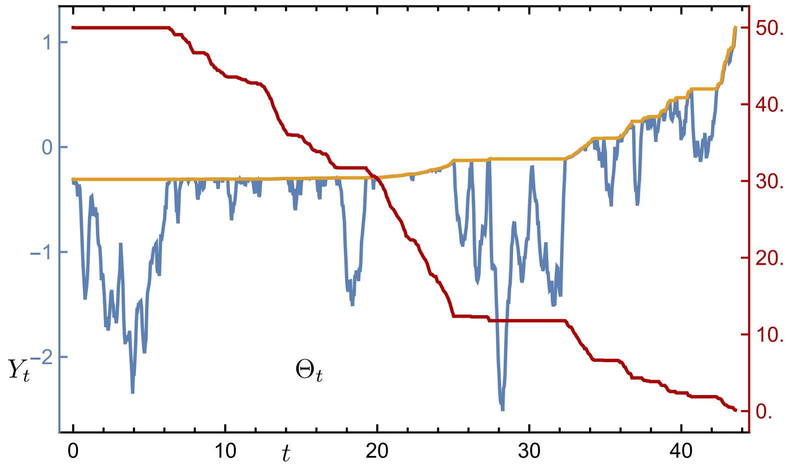


Figure 3.1: Sample path of impact Y_t (blue), asset position Θ_t (red, decreasing) and reflecting boundary $\mathrm{y}(\Theta_t)$ (orange, increasing) for optimally liquidating $\Theta_0 = 50$ assets (after an initial block trade Δ) for $f(\cdot) = \exp(\cdot)$.

displacement of the boundary. Thus, when impact and fundamental price are positively correlated ($\rho > 0$), it is optimal to trade slower if fundamental price volatility is larger, since the wait region increases.

Remark 3.2.4. The optimal liquidation problem with deterministic impact dynamics ($\hat{\sigma} = 0$) is solved in [BBF17a, Thm. 3.4] and characterized by an optimal boundary function y^0. Assumption 3.1.2 implies the model assumptions [BBF17a, Assumption 3.2] of that theorem. Using the asymptotic expansion [Leb72, eq. (10.6.6)] of Hermite functions, straightforward calculations show uniform convergence $\|\mathrm{y}^{\hat{\sigma}} - \mathrm{y}^0\|_\infty \to 0$ of the boundaries as $\hat{\sigma} \searrow 0$, for $\mathrm{y}^{\hat{\sigma}}$ solving the ODE (3.10).

3.3 Reformulation as a calculus of variations problem

In this section we will recast the free boundary problem of the variational inequality (3.9) as a (nonstandard, at first) calculus of variations problem. To sketch the idea, suppose that the large trader has to liquidate $\Theta_0 \geq 0$ shares and that (Y_0, Θ_0) is already on the free boundary between sell and wait regions (after an initial jump or waiting). Let $\mathrm{y} : [0, \Theta_0] \to \mathbb{R}$ be a C^1 function with $\mathrm{y}(\Theta_0) = Y_0$ and $\mathrm{y}' < 0$ (we expect the optimal boundary to be such). To find the optimal boundary curve y, we will optimize expected proceeds over the set of y-reflected strategies $A := A^{\mathrm{refl}}(\mathrm{y}, \Theta_0)$ from

Definition 3.3.1. *Let (Y, A) be the (unique) pair of continuous adapted processes with non-decreasing A such that $Y_t \leq \mathrm{y}(\Theta_0 - A_t)$ and*

$$
\begin{aligned}
\mathrm{d}Y_t &= -\beta Y_t \,\mathrm{d}t + \hat{\sigma}\,\mathrm{d}B_t - \mathrm{d}A_t\,, && Y_0 = \mathrm{y}(\Theta_0)\,,\\
\mathrm{d}A_t &= \mathbb{1}_{\{Y_t = \mathrm{y}(\Theta_0 - A_t)\}}\,\mathrm{d}A_t\,, && A_0 = 0\,,
\end{aligned}
$$

on $[\![0,\tau]\!]$ *for* $\tau := \inf\{t \geq 0 : A_t = \Theta_0\}$. *We call* $A^{\mathit{refl}}(\mathrm{y},\Theta_0) := A$ *a* y-reflected strategy.

Remark 3.3.2. Existence and uniqueness of a strong solution (Y,A) follows from (a careful extension of) classical results, cf. [DI93], by considering the pair (Y,A) as a (degenerate) diffusion in $\mathbb{R}^2$ with oblique direction of reflection $(-1,+1)$ at a smooth boundary. Considered as a one-dimensional diffusion, the process Y is reflected at a boundary that moves with its local time A. In this sense, we call the reflection *elastic*.

Viewing Y as a diffusion with reflection at y, we can rewrite expected proceeds from A as a deterministic functional of y, see (3.19) below, whose maximizer should describe the optimal strategy. For this step we rely crucially on a representation for the Laplace transform of the inverse local time of reflected diffusions from Theorem 3.3.3. Since the integrand of (3.19) depends on the whole path y, a reparametrization is necessary to obtain a tractable calculus of variations problem (3.21) – (3.22).

Let τ_{Θ_0} be the stopping time when $A = \Theta_0$. For the continuous y-reflected strategy A with proceeds $L := L(\mathrm{y}(\Theta_0); A)$, we have by [DM82, Theorem 57] for any $T \in [0,\infty)$,

$$\begin{aligned}\mathbb{E}[L_T] &= \mathbb{E}\Big[\int_0^{\tau_{\Theta_0}\wedge T} f(Y_t)e^{-\delta t}\mathcal{E}(\sigma W)_t \,\mathrm{d}A_t\Big]\\ &= \mathbb{E}\Big[\mathcal{E}(\sigma W)_T\int_0^{\tau_{\Theta_0}\wedge T} f(Y_t)e^{-\delta t}\,\mathrm{d}A_t\Big].\end{aligned}$$

For fixed T, let $\mathbb{Q}$ be the measure given by $\mathrm{d}\mathbb{Q}/\mathrm{d}\mathbb{P} = \mathcal{E}(\sigma W)_T$ on $\mathcal{F}_T$. Then

$$\mathbb{E}[L_T] = \mathbb{E}^{\mathbb{Q}}\Big[\int_0^{\tau_{\Theta_0}\wedge T} f(Y_t)e^{-\delta t}\,\mathrm{d}A_t\Big]. \tag{3.12}$$

Girsanov's theorem gives that the process $\widetilde{B}_t := B_t - [B,\sigma W]_t = B_t - \sigma\rho t$ is a Brownian motion under $\mathbb{Q}$. Therefore, we have under $\mathbb{Q}$

$$\mathrm{d}Y_t = (\sigma\rho\hat{\sigma} - \beta Y_t)\,\mathrm{d}t + \hat{\sigma}\,\mathrm{d}\widetilde{B}_t - \,\mathrm{d}A_t\,,$$

i.e. the impact process Y is a (reflected) Ornstein-Uhlenbeck process with shifted (non-zero) mean reversion level, and A is its local time on the boundary. We cannot directly pass to the limit $T \to \infty$ in (3.12) because the measure change $\mathbb{Q}$ depends on T. However, note that the right-hand side of (3.12) depends only on the law of the reflected diffusion (Y,A) under the measure $\mathbb{Q}$. That is why we consider the reflected diffusion (X,A^X) with the following dynamics under $\mathbb{P}$: for $g(a) := \mathrm{y}(\Theta_0 - a)$ let

$$\mathrm{d}X_t = (\sigma\rho\hat{\sigma} - \beta X_t)\,\mathrm{d}t + \hat{\sigma}\,\mathrm{d}B_t - \,\mathrm{d}A^X_t, \qquad X_0 = g(0)\,, \tag{3.13}$$

$$\mathrm{d}A^X_t = \mathbb{1}_{\{X_t = g(A^X_t)\}}\,\mathrm{d}A^X_t, \qquad A^X_0 = 0\,, \tag{3.14}$$

$$\tau^X_\ell := \inf\{t > 0 : A^X_t > \ell \text{ or } A^X_t = \Theta_0\}\,, \tag{3.15}$$

such that in addition $X_t \leq g(A^X_t)$, on $[\![0,\tau^X_{\Theta_0}]\!]$. Existence and uniqueness of a strong solution (X,A^X) until $\tau^X_{\Theta_0}$ follows as in Remark 3.3.2.

Now, by (3.12) we have $\mathbb{E}[L_T] = \mathbb{E}[\int_0^{\tau^X_{\Theta_0}\wedge T} f(X_t)e^{-\delta t}\,\mathrm{d}A^X_t]$, which gives for $T\to\infty$ by monotone convergence on both sides

$$\begin{aligned}\mathbb{E}[L_\infty] &= \mathbb{E}\Big[\int_0^{\tau^X_{\Theta_0}} f(X_t)e^{-\delta t}\,\mathrm{d}A^X_t\Big] = \mathbb{E}\Big[\int_0^{\tau^X_{\Theta_0}} f\big(g(A^X_t)\big)e^{-\delta t}\,\mathrm{d}A^X_t\Big]\\ &= \mathbb{E}\Big[\int_0^{\Theta_0} f\big(g(\ell)\big)e^{-\delta\tau^X_\ell}\,\mathrm{d}\ell\Big] = \int_0^{\Theta_0} f\big(g(\ell)\big)\mathbb{E}\big[e^{-\delta\tau^X_\ell}\big]\,\mathrm{d}\ell\,, \end{aligned} \tag{3.16}$$

using (3.14). To express the latter as a functional of the free boundary only, we need

Theorem 3.3.3. *The Laplace transform of τ^X_ℓ from (3.13)–(3.15) for $\Theta_0=\theta$ is*

$$\mathbb{E}\big[e^{-\delta\tau^X_\ell}\big] = \exp\Big(\int_{\theta-\ell}^{\theta}\big(\mathrm{y}'(x)-1\big)\frac{\Phi'_\delta(\mathrm{y}(x))}{\Phi_\delta(\mathrm{y}(x))}\,\mathrm{d}x\Big)\quad \textit{for } \ell<\theta. \tag{3.17}$$

Proof. We will identify the Laplace transform by calculating the terms in (3.16) at first for f being replaced with arbitrary test functions φ, and then using ideas from calculus of variations. To identify $q(y,\theta) := \mathbb{E}[\int_0^T e^{-\delta t}\varphi(X_t)\,\mathrm{d}A^X_t]$ for continuous functions $\varphi:\mathbb{R}\to[0,\infty)$ with $X_0=y\le \mathrm{y}(\theta)$, $\Theta_0=\theta$ and $T:=\tau^X_\theta$, it suffices to construct q such that

$$M_t := \int_0^t e^{-\delta u}\varphi(X_u)\,\mathrm{d}A^X_u + e^{-\delta t}q\big(X_t,\theta-A^X_t\big)$$

is a martingale on $[\![0,T]\!]$ with $e^{-\delta t}q(X_t,\theta-A^X_t)\to 0$ in L^1 as $t\to T$. Consider the state space $\mathcal{I} := \{(y,\theta): y<\mathrm{y}(\theta)\}$. To check the martingale property, assuming that we have $q\in C^{2,1}(\mathcal{I})\cap C^{1,1}(\overline{\mathcal{I}})$, Itô's formula yields (similarly to (3.8)) that $q_y+q_\theta=\varphi$ on $\partial\mathcal{I}$ and $\mathcal{L}q(y,\theta)=0$ in $\mathcal{I}$. Moreover, for q increasing in y we have $q(y,\theta)=\Phi(y)C(\theta)$ with $\Phi=\Phi_\delta$ from (3.6) and some function $C\in C^1$. Let $H(\theta):=q(\mathrm{y}(\theta),\theta)$. The condition $q_y+q_\theta=\varphi$ leads to

$$H'(\theta) = \Phi'(\mathrm{y}(\theta))C(\theta)\mathrm{y}'(\theta) + \big(\varphi(\mathrm{y}(\theta)) - \Phi'(\mathrm{y}(\theta))C(\theta)\big) = A(\theta)H(\theta)+B(\theta)$$

where $A(\theta) := \big(\mathrm{y}'(\theta)-1\big)\Phi'(\mathrm{y}(\theta))/\Phi(\mathrm{y}(\theta))$ and $B(\theta):=\varphi(\mathrm{y}(\theta))$. Solving this ODE for H gives (since $H(0)=0$)

$$H(\theta) = \int_0^\theta \varphi\big(\mathrm{y}(z)\big)\exp\Big(\int_z^\theta\big(\mathrm{y}'(x)-1\big)\frac{\Phi'_\delta(\mathrm{y}(x))}{\Phi_\delta(\mathrm{y}(x))}\,\mathrm{d}x\Big)\,\mathrm{d}z,$$

which yields the candidate $q(y,\theta)=\Phi(y)H(\theta)/\Phi(\mathrm{y}(\theta))$. It is straightforward to check $q\in C^{2,1}(\mathcal{I})\cap C^{1,1}(\overline{\mathcal{I}})$ and $q_y+q_\theta=\varphi$ on $\partial\mathcal{I}$, giving that M is a martingale, using boundedness of $q_y(X,\theta-A^X)$ on $[\![0,T]\!]$. By monotonicity of q in y, hence $q(y,\theta)\le H(\theta)$, we obtain $e^{-\delta t}q(X_t,\theta-A^X_t)\to 0$ in L^1 as $t\to T$ via dominated convergence, so as in

(3.16) we find

$$\int_0^\theta \varphi(\mathbb{y}(z))\underbrace{\left(\mathbb{E}[e^{-\delta\tau^X_{\theta-z}}] - \exp\Big(\int_z^\theta \big(\mathbb{y}'(x)-1\big)\frac{\Phi_\delta'(\mathbb{y}(x))}{\Phi_\delta(\mathbb{y}(x))}\,\mathrm{d}x\Big)\right)}_{=:\Delta(z)}\mathrm{d}z = 0. \tag{3.18}$$

Note that $z \mapsto \mathbb{E}[\exp(-\delta\tau^X_{\theta-z})]$ is left-continuous. Hence, if $\Delta(z_1) > 0$ for some $z_1 \in (0,\theta]$, there exists $z_0 < z_1$ such that $\Delta > 0$ on (z_0, z_1). Since $\mathbb{y}$ is bijective (recall that $\mathbb{y}' < 0$), we can find a continuous function φ with $\varphi \circ \mathbb{y} > 0$ inside (z_0, z_1) and $\varphi \circ \mathbb{y} = 0$ outside (z_0, z_1), which yields $\int_0^\theta \varphi(\mathbb{y}(z))\Delta(z)\,\mathrm{d}z > 0$, contradicting (3.18). Similarly, $\Delta(z_1) < 0$ also leads to a contradiction. Therefore $\Delta = 0$ on $(0,\theta]$. □

Remark 3.3.4. Let us note that Theorem 3.3.3 generalizes to general (regular) diffusions reflected at increasing boundaries by taking Φ_δ to be the increasing non-negative δ-eigenfunction of the generator of the diffusion. Indeed, the proof would not change.

Remark 3.3.5. The representation (3.17) can be also derived by a probabilistic approximation argument. Indeed, consider the following natural approximation $(X^\varepsilon, A^\varepsilon)$ of the continuous reflected diffusion (X, A): A^ε increases by ε (by a jump) whenever X^ε hits the boundary. Thus, one obtains a sequence of processes for which the Laplace transform of the inverse local time of A^ε can be expressed in terms of excursion lengths whose Laplace transforms are known. One can then show that $(X^\varepsilon, A^\varepsilon)$ converges in distribution to (X, A), as the jump size ε goes to zero, and thus derive (3.17) in the limit, see [BBF18a] for details. It can be moreover shown that the proceeds from A^ε satisfy $\mathbb{E}[L_\infty(A)] = \mathbb{E}[L_\infty(A^\varepsilon)] + \mathcal{O}(\varepsilon)$, see [BBF18a, Remark 3.4]. Thus, the simple process A^ε approximates the reflection local time A up to first order, an appealing property from practical point of view because A is typically singular (as a local-time process) and so implementing the trading strategy A can be approximated by implementing A^ε.

Using Theorem 3.3.3 and (3.16) we derive the following representation for the proceeds from a $\mathbb{y}$-reflected strategy in terms of the boundary:

$$\mathbb{E}[L_\infty] = \int_0^{\Theta_0} f\big(g(\ell)\big)\exp\Big(-\int_0^\ell \big(g'(a)+1\big)\frac{\Phi_\delta'(g(a))}{\Phi_\delta(g(a))}\,\mathrm{d}a\Big)\,\mathrm{d}\ell\,. \tag{3.19}$$

Since the $\mathrm{d}\ell$-integrand in (3.19) depends on the whole path of g, classical calculus of variations methods are not directly available. Since by definition $g(a) = \mathbb{y}(\Theta_0 - a)$ we get with $\mathbb{r}(\ell) := \int_0^\ell (1-\mathbb{y}'(x))\frac{\Phi'(\mathbb{y}(x))}{\Phi(\mathbb{y}(x))}\,\mathrm{d}x$ that

$$\mathbb{E}[L_\infty] = e^{-\mathbb{r}(\Theta_0)}\int_0^{\Theta_0} f\big(\mathbb{y}(\ell)\big)e^{\mathbb{r}(\ell)}\,\mathrm{d}\ell. \tag{3.20}$$

Since $\Phi', \Phi > 0$ and $\mathbb{y}' < 0$, the function $\mathbb{r}$ in strictly increasing and thus has an inverse $\mathbb{r}^{-1}$. Fixing $R := \mathbb{r}(\Theta_0)$ and setting $w(r) := \mathbb{y}(\mathbb{r}^{-1}(r))$, we find

$$\mathbb{r}^{-1}(r) = \int_0^r \left(w'(z) + \frac{\Phi(w(z))}{\Phi'(w(z))}\right)\mathrm{d}z.$$

Hence, by the reparametrization $\mathbb{y}(\theta) = w(\mathbb{r}(\theta))$, finding a maximizing function $\mathbb{y}$ for (3.20) reduces to the problem of finding a function w which maximizes

$$J(w) := \int_0^R f\big(w(r)\big)e^{-(R-r)}\Big(w'(r) + \frac{\Phi(w(r))}{\Phi'(w(r))}\Big)\,\mathrm{d}r \quad (= \mathbb{E}[L_\infty]) \tag{3.21}$$

$$\text{subject to the condition } K(w) := \int_0^R \Big(w'(r) + \frac{\Phi(w(r))}{\Phi'(w(r))}\Big)\,\mathrm{d}r = \Theta_0\,. \tag{3.22}$$

3.4 Solving the calculus of variations problem

In this section, we solve (locally) the calculus of variations problem of maximizing (3.21) subject to (3.22) by employing necessary and sufficient conditions on the first and second variation of the functionals involved. We obtain the candidate free boundary function $\mathbb{y}(\theta)$, see equations (3.28) and (3.29), and show its local optimality in Lemma 3.4.4. We then relate our results on the calculus of variations problem to the initial optimal execution problem in Theorem 3.4.6. This will be crucial later for Section 3.5 to verify the desired inequality in the sell region, presented in Lemma 3.5.7.

A maximizer w of the isoperimetric problem (3.21) – (3.22) also maximizes $J + mK$ for some constant $m := m(R)$ that is the Lagrange multiplier, cf. [GF00, Theorem 2.12.1]. Considering perturbations $w(r) + h(r)$ of w with $h(0) = h(R) = 0$, a necessary condition for an extremum w of a functional $J + mK$ is that its first variation $\mathrm{D}(J + mK)$ vanishes at w, see [GF00, Thm. 1.3.2]. Integration by parts yields the Euler-Lagrange equation

$$0 = F_w - \frac{\mathrm{d}}{\mathrm{d}r}F_{w'} + \Big(G_w - \frac{\mathrm{d}}{\mathrm{d}r}G_{w'}\Big)m\,, \tag{3.23}$$

with $G(r, w, w') := w' + \Phi(w)/\Phi'(w)$ and $F(r, w, w') := f(w)e^{-(R-r)}G(r, w, w')$, the integrands of K and J, respectively.

Since we assumed to start on the (yet unknown) boundary, one side is fixed, we have $w(R) = \mathbb{y}(\Theta_0)$. But the other end $w(0)$ is free. Thus, integration by parts of $\mathrm{D}(J + mK)$ with perturbations $w(r) + h(r)$ of w where $h(0) \neq 0$ imposes as an additional condition for $\mathrm{D}(J + mK)$ to vanish that

$$0 = \big(F_{w'} + mG_{w'}\big)\big|_{r=0}\,.$$

This *natural boundary condition* (cf. [GF00, Sect. 1.6]) yields the Lagrange multiplier $m(R) = -f(y_0)e^{-R}$ for $y_0 := \mathbb{y}(0) = w(0)$. After multiplication with $e^R\Phi'(w)^2$, equation (3.23) simplifies to

$$e^r\Phi(w)\big(f'(w)\Phi'(w) - f(w)\Phi''(w)\big) = f(y_0)\big(\Phi'(w)^2 - \Phi(w)\Phi''(w)\big)\,. \tag{3.24}$$

Inserting $r = 0$ gives a condition for y_0, namely

$$f'(y_0)\Phi(y_0) = f(y_0)\Phi'(y_0).$$

Assumption C6 guarantees existence and C2 uniqueness of y_0. On the other hand, differentiating both sides of (3.24) with respect to r gives the ODE for w

$$\begin{aligned} 0 = \big(e^r(f'\Phi' - f\Phi'')\Phi' + e^r(f''\Phi' - f\Phi''')\Phi - f(y_0)(\Phi'\Phi'' - \Phi\Phi''')\big)w' \\ + e^r(f'\Phi' - f\Phi'')\Phi, \end{aligned} \tag{3.25}$$

where $f = f(w(r))$, $f' = f'(w(r))$, $\Phi = \Phi(w(r))$, etc.

Both sides in the above equality (3.24) are negative on the boundary $w(r)$, due to

Lemma 3.4.1. *The positive, increasing eigenfunctions $\Phi = \Phi_\delta$ corresponding to the eigenvalue $\delta > 0$ of the generator of an Ornstein-Uhlenbeck process satisfy*

$$\big(\Phi^{(n)}(x)\big)^2 < \Phi^{(n-1)}(x)\Phi^{(n+1)}(x)$$

for all $x \in \mathbb{R}$ and $n \in \mathbb{N}$. In particular, $(\Phi')^2 < \Phi\Phi''$. Moreover, for $n \in \mathbb{N}$

$$\lim_{x\to-\infty} \Phi^{(n)}(x)/\Phi^{(n-1)}(x) = 0 \quad \text{and} \quad \lim_{x\to+\infty} \Phi^{(n)}(x)/\Phi^{(n-1)}(x) = +\infty.$$

Proof. Since $H'_\nu(x) = 2\nu H_{\nu-1}(x)$ for complex ν (see [Leb72, eq. (10.4.4)]), equation (3.6) implies

$$\Phi_\delta^{(n)}\Phi_\delta^{(n+2)} - \big(\Phi_\delta^{(n+1)}\big)^2 = \big(\Phi_{\delta+n\beta}\Phi''_{\delta+n\beta} - (\Phi'_{\delta+n\beta})^2\big)\frac{2^{2n}}{\hat\sigma^{2n}\beta^n}\prod_{k=0}^{n}(\delta + k\beta)^2\,,$$

so it suffices to prove $(\Phi')^2 < \Phi''\Phi$ for every $\delta, \beta, \sigma, \hat\sigma > 0$ and $\rho \in [-1, 1]$ in (3.6). This is equivalent to showing $(H'_\nu)^2 < H''_\nu H_\nu$ for every $\nu < 0$. Since $\Gamma(-\nu) > 0$ and $H_\nu(x) = \Gamma(-\nu)^{-1}\int_0^\infty e^{-t^2-2xt}t^{-\nu-1}\,\mathrm{d}t$ for $\nu < 0$ (cf. [Leb72, eq. (10.5.2)]), the function $\varphi_x(t) := e^{-t^2-2xt}t^{-\nu-1}$ is the density of an absolutely continuous finite measure μ on $[0,\infty)$. For the probability measure $\tilde{\mathbb{P}}[A] := \mu([0,\infty))^{-1}\mu(A)$ consider two independent random variables $X, Y \sim \tilde{\mathbb{P}}$. By [Kle08, Thm. 6.28], we can exchange differentiation and integration (in the integral representation of H_ν above) to see that $H''_\nu(x)H_\nu(x) - H'_\nu(x)^2 = 4\,\tilde{\mathbb{E}}[X^2 - XY]$. Symmetry gives $2\,\tilde{\mathbb{E}}[X^2 - XY] = \tilde{\mathbb{E}}[(X-Y)^2] \geq 0$. Since X and Y are independent with absolutely continuous distribution, Fubini's theorem yields $\tilde{\mathbb{P}}[X = Y] = 0$, so $\tilde{\mathbb{E}}[(X-Y)^2] > 0$.

The asymptotic behavior of $\Phi^{(n)}/\Phi^{(n-1)}$ follows from [Leb72, eq. (10.6.4)] in the case $x \to -\infty$ and from [Leb72, eq. (10.6.7)] in the case $x \to +\infty$. □

Now (3.24) gives a representation of r given y_0 and w as

$$r = \log\frac{f(y_0)}{\Phi(w)} + \log\frac{\Phi'(w)^2 - \Phi(w)\Phi''(w)}{f'(w)\Phi'(w) - f(w)\Phi''(w)}, \tag{3.26}$$

which we can use to simplify the ODE (3.25) (assuming $w' \neq 0$ everywhere) to

$$\frac{1}{w'} = -\frac{\Phi'}{\Phi} + \frac{f\Phi''' - f''\Phi'}{f'\Phi' - f\Phi''} + \frac{\Phi'\Phi'' - \Phi\Phi'''}{(\Phi')^2 - \Phi\Phi''},$$

reading the right hand side as a function of $w(r)$. With $\mathtt{y}(\theta) = w(\mathtt{r}(\theta))$ and $r := \mathtt{r}(\theta)$, we get $\mathtt{y}'(\theta) = w'(r)\mathtt{r}'(\theta) = w'(r)(1 - \mathtt{y}'(\theta))\Phi'(\mathtt{y}(\theta))/\Phi(\mathtt{y}(\theta))$, which simplifies to

$$\begin{aligned}
\mathtt{y}'(\theta) &= \frac{\Phi'(\mathtt{y})}{\Phi'(\mathtt{y}) + \Phi(\mathtt{y})/w'(r)} \\
&= \frac{1}{\Phi} \frac{((\Phi')^2 - \Phi\Phi'')(f'\Phi' - f\Phi'')}{(\Phi\Phi'' - (\Phi')^2)f'' + (\Phi'\Phi'' - \Phi\Phi''')f' + (\Phi'\Phi''' - (\Phi'')^2)f} \\
&= \frac{M_2(\mathtt{y}(\theta))}{M_1'(\mathtt{y}(\theta))},
\end{aligned} \tag{3.27}$$

$$\text{where} \quad M_1 := \frac{f\Phi' - f'\Phi}{(\Phi')^2 - \Phi\Phi''} \qquad \text{and} \qquad M_2 := \frac{f'\Phi' - f\Phi''}{(\Phi')^2 - \Phi\Phi''}. \tag{3.28}$$

By (3.24) and Lemma 3.4.1 we have $M_2(\mathtt{y}(\theta)) > 0$ for any θ. We get $M_1'(\mathtt{y}(\theta)) < 0$ by

Lemma 3.4.2. *Under Assumption C2, $M_1'(y) < 0$ for all $y \in \mathbb{R}$.*

Proof. Let $G := \Phi'/\Phi$ and $H := \Phi''/\Phi'$. We have $G, G', H, H' > 0$ and $G < H$ by Lemma 3.4.1. With $\lambda(y) = f'(y)/f(y) > 0$, thus $f''/f = \lambda' + \lambda^2$, we get

$$(G')^2\Phi M_1'/f = \lambda' G' + (\lambda^2 - \lambda H)G' + (G^2 - \lambda G)H'.$$

So $M_1'(y) < 0$ if and only if $\lambda'(y)G'(y) < q(\lambda(y))$ where the right-hand side is $q(\lambda) := (H - \lambda)\lambda G' + (\lambda - G)GH'$. The function q is quadratic in λ and takes its minimum in

$$\lambda^* := \frac{HG' + GH'}{2G'} \qquad \text{with value} \qquad q(\lambda^*) = \frac{(HG' + GH')^2}{4G'} - G^2H'.$$

Note also, that $G' = (H - G)G$. We find that

$$\begin{aligned}
4G'\left(\lambda' G' - q(\lambda)\right) &\le 4G'\left(\lambda' G' - q(\lambda^*)\right) < 4G'\left((G')^2 - q(\lambda^*)\right) \\
&= 4(G')^3 - (GH' + G'H)^2 + 4G'G^2H' \\
&= G^2\Big(4G(H - G)^3 - \big(H' + (H - G)H\big)^2 + 4(H - G)GH'\Big) \\
&= -G^2\big(H' + H^2 + 2G^2 - 3GH\big)^2 \le 0,
\end{aligned}$$

using that $\lambda'(y) < G'(y)$, $y \in \mathbb{R}$, by Assumption C2. So $M_1'(y) < 0$ for all $y \in \mathbb{R}$. $\square$

Lemma 3.4.3. *Let f satisfy Assumptions C2, C3 and C6. Then there exists a unique solution $\theta \mapsto \mathtt{y}(\theta)$, $\theta \in [0, \infty)$, of the ODE*

$$\mathtt{y}' = M_2(\mathtt{y})/M_1'(\mathtt{y}), \quad \mathtt{y}(0) = y_0, \tag{3.29}$$

and $\mathtt{y}$ is strictly decreasing to $\lim_{\theta\to\infty} \mathtt{y}(\theta) = y_\infty$ (with y_0 and y_∞ from Assumption C6).

Proof. Since M_2/M_1' is locally Lipschitz by $f \in C^3(\mathbb{R})$, there exists a unique maximal solution $\mathtt{y} : [0, \theta_{\max}) \to \mathbb{R}$ of (3.29). We have $M_2(\mathtt{y}(\theta)) > 0$ and $M_1' < 0$ by Lemma 3.4.2, thus $\mathtt{y}' < 0$. Assume $\theta_{\max} < \infty$, which implies $\lim_{\theta \to \theta_{\max}} \mathtt{y}(\theta) = -\infty$. However, note that $\{(\theta, \mathtt{y}(\theta)) : 0 \le \theta < \theta_{\max}\}$ and $[0, \infty) \times \{y_\infty\}$ are trajectories of the two-dimensional autonomous dynamical system induced by the field $(\theta, y) \mapsto (1, M_2(y)/M_1'(y))$. Since trajectories of autonomous dynamical systems cannot cross, and $y_\infty < y_0$ by Lemma 3.4.1, we must have $y_\infty < \mathtt{y}(\theta)$ for all $\theta \in [0, \theta_{\max})$, which contradicts $\theta_{\max} < \infty$.

Moreover, $\mathtt{y}^{-1}(y) = \int_{y_0}^{y} (M_1'/M_2)(x)\,\mathrm{d}x$ is finite for every $y \in (y_\infty, y_0]$. Since $\theta_{\max} = \infty$, it follows that $\mathtt{y}(\theta) \to y_\infty$ as $\theta \to \infty$. $\square$

By considering the first variation $\mathrm{D}(J + mK)$, we found a candidate boundary function $\mathtt{y}$ in terms of a possible extremum $w : [0, R] \to \mathbb{R}$ of $J + mK$. Calculating the second variation $\mathrm{D}^2(J + mK)$ at w, we find that w is indeed a local maximizer.

Lemma 3.4.4. *The functional* $\hat{J} := J + mK : C^1([0, R]) \to \mathbb{R}$ *defined by* (3.21) – (3.22) *with* $m = -f(y_0)e^{-R}$ *has a strict local maximizer* $w(r) = \mathtt{y}(\mathtt{r}^{-1}(r))$, *with* $\mathtt{y}$ *solving* (3.29), *in the following sense. There exists* $\varepsilon > 0$ *such that for all perturbations* $0 \not\equiv h \in C^1([0, R])$ *with endpoints* $h(0) = h(R) = 0$ *and* $\|h\|_{W^{1,\infty}} = \|h\|_\infty \vee \|h'\|_\infty < \varepsilon$ *it holds*

$$\hat{J}(w + h) < \hat{J}(w).$$

Proof. For a C^1-perturbation $h : [0, R] \to \mathbb{R}$ of w with $h(0) = h(R) = 0$ we have by [GF00, Sect. 5.25, (10) and (11)]

$$\mathrm{D}^2(J + mK)[w; h] = \int_0^R \big(Ph'(r)^2 + Qh(r)^2\big)\,\mathrm{d}r$$

with $P = P(r, w(r), w'(r))$ and $Q = Q(r, w(r), w'(r))$ given by

$$\begin{aligned}
P &= \tfrac{1}{2}\big(F_{w'w'} + mG_{w'w'}\big) = 0,\\
Q &= \frac{1}{2}\Big(F_{ww} + mG_{ww} - \frac{\mathrm{d}}{\mathrm{d}r}\Big(F_{ww'} + mG_{ww'}\Big)\Big)\\
&= \frac{1}{2}e^{-(R-r)}\bigg(\frac{\Phi}{\Phi'}f'' + 2\Big(\frac{\Phi}{\Phi'}\Big)'f' + \Big(\frac{\Phi}{\Phi'}\Big)''f - f'\bigg) + \frac{1}{2}\Big(\frac{\Phi}{\Phi'}\Big)''m,
\end{aligned}$$

with f, Φ and their derivatives being evaluated at $w(r)$ when no argument is mentioned. Differentiating (3.23) with respect to r yields

$$\begin{aligned}
0 &= \frac{\mathrm{d}}{\mathrm{d}r}e^{-(R-r)}\bigg(\frac{\Phi}{\Phi'}f' + \Big(\frac{\Phi}{\Phi'}\Big)'f - f\bigg) + m\,\frac{\mathrm{d}}{\mathrm{d}r}\Big(\frac{\Phi}{\Phi'}\Big)'\\
&= e^{-(R-r)}\bigg(\frac{\Phi}{\Phi'}f' + \Big(\frac{\Phi}{\Phi'}\Big)'f - f\bigg)\\
&\qquad + e^{-(R-r)}\bigg(\frac{\Phi}{\Phi'}f'' + 2\Big(\frac{\Phi}{\Phi'}\Big)'f' + \Big(\frac{\Phi}{\Phi'}\Big)''f - f'\bigg)w' + \Big(\frac{\Phi}{\Phi'}\Big)''m\,w'
\end{aligned}$$

$$
\begin{aligned}
&= e^{-(R-r)}\left(\frac{\Phi}{\Phi'}f' + \left(\frac{\Phi}{\Phi'}\right)' f - f\right) + 2Qw' \\
&= e^{-(R-r)}\frac{\Phi}{(\Phi')^2}\left(f'\Phi' - f\Phi''\right) + 2Qw' \,. \qquad (3.30)
\end{aligned}
$$

By equation (3.24) and Lemma 3.4.1, the first summand in (3.30) is negative along $w(r)$. Since $w(r) = \mathtt{y}(\mathtt{r}^{-1}(r))$ and $\mathtt{r}^{-1}$ is strictly increasing, we have $w' < 0$ by Lemma 3.4.3. So $Q(r, w(r), w'(r)) < -\kappa < 0$ on $[0, R]$ by (3.30) for some constant $\kappa = \kappa_R$, giving that the second variation is negative definite at w, i.e. for $h \not\equiv 0$,

$$
\mathrm{D}^2(J+mK)[w;h] = \int_0^R Q(r, w(r), w'(r))h(r)^2 \,\mathrm{d}r < -\kappa \int_0^R h(r)^2 \,\mathrm{d}r < 0\,. \qquad (3.31)
$$

To shorten notation, let $\hat{F} := F + mG$, so $\hat{J} := J + mK = \int_0^R \hat{F}\,\mathrm{d}r$. Unless the arguments are explicitly written, take $\hat{F} = \hat{F}(r, w(r), w'(r))$. Taylor's theorem gives $\hat{J}(w+h) - \hat{J}(w) = \mathrm{D}\hat{J}[w;h] + \mathrm{D}^2\hat{J}[w;h] + \mathcal{E}(h)$ with first variation $\mathrm{D}\hat{J}[w;h] = 0$ by (3.23), second variation $\mathrm{D}^2\hat{J}[w;h] = \int_0^R Qh^2\,\mathrm{d}r < 0$ by (3.31) and remainder

$$
\mathcal{E}(h) = \int_0^R \left(\sum_{|\boldsymbol{\alpha}|=3} \partial^{\boldsymbol{\alpha}} \hat{F}\big(r, \boldsymbol{w} + \xi_r \boldsymbol{h}\big) \frac{\boldsymbol{h}^{\boldsymbol{\alpha}}}{\boldsymbol{\alpha}!} \right) \mathrm{d}r
$$

for some $\xi_r \in [0,1]$, with $\boldsymbol{w} = (w(r), w'(r))^\top$, $\boldsymbol{h} = (h(r), h'(r))^\top$ and multi-index $\boldsymbol{\alpha} \in \mathbb{N}_0^2$, considering $\hat{F}(r,\cdot)$ as an function on $\mathbb{R}^2$. Since $\hat{F}$ is affine in w' we get

$$
\mathcal{E}(h) = \int_0^R \Big(\frac{1}{6}\hat{F}_{www}(r, \boldsymbol{w} + \xi_r \boldsymbol{h})h + \frac{1}{2}\hat{F}_{www'}(r, \boldsymbol{w} + \xi_r \boldsymbol{h})h' \Big) h^2 \,\mathrm{d}r =: \int_0^R A h^2 \,\mathrm{d}r
$$

Note that by compactness of $[0, R]$ we have uniform convergence

$$
\sup_{r\in[0,R]} \sup_{\xi\in[0,1]} \big| A\big(h(r), h'(r), w(r), w'(r), \xi, r\big) \big| \to 0
$$

as $\|h\|_{W^{1,\infty}} \to 0$. Now choose $\varepsilon > 0$ small enough such that

$$
\big| A\big(h(r), h'(r), w(r), w'(r), \xi, r\big) \big| < \kappa/2
$$

for all $r \in [0, R]$, $\xi \in [0,1]$ and h with $\|h\|_{W^{1,\infty}} < \varepsilon$. Hence, with $h \not\equiv 0$

$$
\hat{J}(w+h) - \hat{J}(w) = \int_0^R (Q + A)h^2 \,\mathrm{d}r < -\frac{\kappa}{2} \int_0^R h^2 \,\mathrm{d}r < 0\,. \qquad \square
$$

Note that the definition $w(r) := \mathtt{y}(\mathtt{r}^{-1}(r))$ does not depend on the interval boundary R. Hence the optimizer w over $[0, R]$ from Lemma 3.4.4 is optimal for all $R > 0$. We can calculate the value $J(w)$ of our optimizer explicitly.

Lemma 3.4.5. *For the optimal w from Lemma 3.4.4 we have*

$$J(w) = (\Phi M_1)(\mathrm{y}(\Theta_0)) = (\Phi M_1)(w(R)).$$

Proof. By direct calculation we have $fM_1'/(\Phi M_2^2) = ((f\Phi' - f'\Phi)/(f'\Phi' - f\Phi''))'$. Moreover, (3.24) gives $e^r = f(y_0)/(\Phi M_2)(w(r))$. With $r = \mathrm{r}(\ell)$ and using (3.29), we get from (3.20) that

$$\begin{aligned}
J(w) &= e^{-\mathrm{r}(\Theta_0)} \int_0^{\Theta_0} f(\mathrm{y}(\ell)) e^{\mathrm{r}(\ell)} \, \mathrm{d}\ell \\
&= (\Phi M_2)(\mathrm{y}(\Theta_0)) \int_0^{\Theta_0} \Big(\frac{f}{\Phi M_2} \Big) (\mathrm{y}(\ell)) \, \mathrm{d}\ell \\
&= (\Phi M_2)(\mathrm{y}(\Theta_0)) \int_{y_0}^{\mathrm{y}(\Theta_0)} \Big(\frac{f M_1'}{\Phi M_2^2} \Big) (x) \, \mathrm{d}x \\
&= (\Phi M_2)(\mathrm{y}(\Theta_0)) \left[\frac{f\Phi' - f'\Phi}{f'\Phi' - f\Phi''} \right]_{y_0}^{\mathrm{y}(\Theta_0)} \\
&= (\Phi M_1)(\mathrm{y}(\Theta_0)).
\end{aligned}$$

□

Now we can translate the results obtained so far back to the state space of impact and asset position. The following theorem will be crucial for our analysis in the verification arguments in Section 3.5.

Theorem 3.4.6. *The function $\mathrm{y} : [0, \infty) \to \mathbb{R}$ defined by equation (3.29) is a (one-sided) local maximizer of $\mathbb{E}[L_\infty(A^{refl}(\mathrm{y}, \Theta_0))]$ in the sense that, for every $\theta > 0$ there exists $\varepsilon > 0$ such that for any decreasing $\tilde{\mathrm{y}} \in C^1([0, \infty))$ with $\mathrm{y}(\cdot) \le \tilde{\mathrm{y}}(\cdot) \le y_0$, $\mathrm{y} = \tilde{\mathrm{y}}$ on $[\theta, \infty)$ and $0 < \|\mathrm{y} - \tilde{\mathrm{y}}\|_{W^{1,\infty}} < \varepsilon$, it holds*

$$\mathbb{E}\big[L_\infty\big(A^{refl}(\mathrm{y}, \theta)\big)\big] > \mathbb{E}\big[L_\infty\big(A^{refl}(\tilde{\mathrm{y}}, \theta)\big)\big].$$

Proof. For sake of clarity, we write $J = J_R$ and $K = K_R$ to emphasize the dependence of the functionals J, K on R. Call $w(r)$ the parametrization of y and $\tilde{w}(r)$ the parametrization of $\tilde{\mathrm{y}}$.

Fix $\theta > 0$ and choose $R, \hat{R}, \hat{\theta}$ such that $\mathrm{y}(\theta) = w(R)$, $\tilde{\mathrm{y}}(\theta) = \tilde{w}(\hat{R})$ and $w(\hat{R}) = \mathrm{y}(\hat{\theta})$. So $R := \mathrm{r}_{\mathrm{y}}(\theta)$, $\hat{R} := \mathrm{r}_{\tilde{\mathrm{y}}}(\theta) = \int_0^\theta \frac{\Phi'}{\Phi}(\tilde{\mathrm{y}}(x)) \, \mathrm{d}x + \int_{\tilde{\mathrm{y}}(\theta)}^{\tilde{\mathrm{y}}(0)} \frac{\Phi'}{\Phi}(u) \, \mathrm{d}u$ and $\hat{\theta} := \mathrm{r}_{\mathrm{y}}^{-1}(\hat{R})$. By $\mathrm{y} \not\equiv \tilde{\mathrm{y}}$, $\mathrm{y}(\cdot) \le \tilde{\mathrm{y}}(\cdot)$ with equality outside $(0, \theta)$ and monotonicity of Φ'/Φ, we have $\hat{R} > R$ and thus $\hat{\theta} > \theta$.

Now, $K_{\hat{R}}(w) = \hat{\theta}$ and $K_{\hat{R}}(\tilde{w}) = \theta$. Moreover, $J_r(w) = (\Phi M_1)(w(r))$ by Lemma 3.4.5.

So if $\|w - \tilde{w}\|_{W^{1,\infty}}$ is small enough, by Lemma 3.4.4 we get

$$\begin{aligned}
J_R(w) &= \big(\Phi M_1\big)\big(w(R)\big) - \big(\Phi M_1\big)\big(w(\hat{R})\big) + J_{\hat{R}}(w) \\
&= \big(\Phi M_1\big)\big(w(R)\big) - \big(\Phi M_1\big)\big(w(\hat{R})\big) + e^{-\hat{R}} f(y_0)\hat{\theta} + \big(J_{\hat{R}} - e^{-\hat{R}} f(y_0) K_{\hat{R}}\big)(w) \\
&> \big(\Phi M_1\big)\big(w(R)\big) - \big(\Phi M_1\big)\big(w(\hat{R})\big) + e^{-\hat{R}} f(y_0)\hat{\theta} + \big(J_{\hat{R}} - e^{-\hat{R}} f(y_0) K_{\hat{R}}\big)(\tilde{w}) \\
&= \big(\Phi M_1\big)\big(\mathrm{y}(\hat{\theta} - \eta)\big) - \big(\Phi M_1\big)\big(\mathrm{y}(\hat{\theta})\big) + e^{-\hat{R}} f(y_0)\eta + J_{\hat{R}}(\tilde{w}) \\
&=: \Psi(\eta) + J_{\hat{R}}(\tilde{w})\,.
\end{aligned}$$

where $\eta := \hat{\theta} - \theta > 0$. By (3.26) we get $e^{-\hat{R}} f(y_0) = (\Phi M_2)(\mathrm{y}(\hat{\theta}))$. With (3.27) follows

$$\begin{aligned}
\Psi'(\eta) &= -\left((\Phi M_1)' \frac{M_2}{M_1'}\right)\big(\mathrm{y}(\hat{\theta} - \eta)\big) + \big(\Phi M_2\big)\big(\mathrm{y}(\hat{\theta})\big) \\
&= -\left(\frac{\Phi' M_1 M_2}{M_1'} + \Phi M_2\right)\big(\mathrm{y}(\hat{\theta} - \eta)\big) + \big(\Phi M_2\big)\big(\mathrm{y}(\hat{\theta})\big).
\end{aligned}$$

Hence $\Psi'(0) = -(\Phi' M_1 M_2 / M_1')(\mathrm{y}(\hat{\theta}))$. Since $M_1 > 0$ on $(-\infty, y_0)$, $M_2 > 0$ on $(y_\infty, y_0]$, $M_1' < 0$ by Lemma 3.4.2 and $\Phi' > 0$, it follows $\Psi'(0) > 0$. So $\Psi(\eta) > 0$ for $\eta > 0$ small enough. Therefore we have by (3.21)

$$\mathbb{E}\big[L_\infty\big(A^{\mathrm{refl}}(\mathrm{y}, \theta)\big)\big] = J_R(w) > J_{\hat{R}}(\tilde{w}) = \mathbb{E}\big[L_\infty\big(A^{\mathrm{refl}}(\tilde{\mathrm{y}}, \theta)\big)\big].$$

The bounds on η and $\|w - \tilde{w}\|_{W^{1,\infty}}$ are satisfied for small enough $\varepsilon > 0$, because $(\mathrm{y}, \ell) \mapsto \mathrm{r}_{\mathrm{y}}(\ell)$ and $(\mathrm{y}, \ell) \mapsto \mathrm{r}_{\mathrm{y}}^{-1}(\ell)$ are continuous in $W^{1,\infty} \times \mathbb{R}$, so $\|w - \tilde{w}\|_{W^{1,\infty}} \to 0$, $\hat{R} \to R$ and $\hat{\theta} \to \theta$ as $\varepsilon \to 0$. □

3.5 Constructing the value function and verification

In this section, we construct a candidate for the value function and verify the variational inequality (3.9) in Lemmas 3.5.6 and 3.5.7, relying on results from the previous sections. This will be sufficient to conclude the proof of our main result, Theorem 3.2.1.

Having defined a candidate boundary via the ODE (3.29) to separate the sell and wait regions $\mathcal{S}$ and $\mathcal{W}$, we will now construct a solution V of the variational inequality (3.9) that will give the value function of the optimal liquidation problem. As a direct consequence of Lemma 3.4.5, we get its value along the boundary

$$V_{\mathrm{bdry}}(\theta) := V\big(\mathrm{y}(\theta), \theta\big) = \Phi\big(\mathrm{y}(\theta)\big) M_1\big(y(\theta)\big)\,. \tag{3.32}$$

Inside the wait region $\mathcal{W}$, which we assume is to the left of the boundary, we require $V = V^{\mathcal{W}}$ to satisfy $\frac{\hat{\sigma}^2}{2} V_{yy} + (\sigma\rho\hat{\sigma} - \beta y) V_y = \delta V$. Note that $V^{\mathcal{W}}$ solves the same ODE in y as Φ. Since V should be also monotonically increasing, the only possibility is that $V^{\mathcal{W}}(y, \theta) = C(\theta)\Phi(y)$ for some increasing function $C : [0, \infty) \to [0, \infty)$. Using the boundary condition $V^{\mathcal{W}}(\mathrm{y}(\theta), \theta) = V_{\mathrm{bdry}}(\theta)$, in light of equation (3.32) we then have

$$V^{\mathcal{W}}(y, \theta) := \Phi(y) C(\theta) \tag{3.33}$$

for $y \leq \mathrm{y}(\theta)$ and $\theta \geq 0$, where $C(\theta) := M_1(\mathrm{y}(\theta))$. On the other hand, in the sell region we require for $V = V^{\mathcal{S}}$ to satisfy $f = V^{\mathcal{S}}_y + V^{\mathcal{S}}_\theta$. We divide $\mathcal{S}$ in two parts:

$$\begin{aligned}
\mathcal{S}_1 &:= \{(y, \theta) \in \mathbb{R} \times (0, \infty) : \mathrm{y}(\theta) < y < y_0 + \theta\}, \\
\mathcal{S}_2 &:= \{(y, \theta) \in \mathbb{R} \times (0, \infty) : y_0 + \theta < y\}.
\end{aligned}$$

Let $\Delta := \Delta(y, \theta) \geq 0$ denote the $\|\cdot\|_\infty$-distance of a point $(y, \theta) \in \overline{\mathcal{S}}$ to the boundary $\partial \mathcal{S}$ in direction $(-1, -1)$. This means in $\overline{\mathcal{S}}_1$ (but not in $\mathcal{S}_2$) that

$$\mathrm{y}(\theta - \Delta) = y - \Delta. \tag{3.34}$$

Inside $\overline{\mathcal{S}}_1$, we need to have

$$V^{\mathcal{S}_1}(y, \theta) := V^{\mathcal{W}}(y - \Delta, \theta - \Delta) + \int_{y-\Delta}^{y} f(x) \, \mathrm{d}x, \tag{3.35}$$

since $V^{\mathcal{S}_1}_y + V^{\mathcal{S}_1}_\theta = f$ in $\overline{\mathcal{S}}$ and $V^{\mathcal{S}_1}(\mathrm{y}(\theta), \theta) = V^{\mathcal{W}}(\mathrm{y}(\theta), \theta)$. Similarly, in $\overline{\mathcal{S}}_2$,

$$V^{\mathcal{S}_2}(y, \theta) := \int_{y-\theta}^{y} f(x) \, \mathrm{d}x. \tag{3.36}$$

To wrap up, the candidate value function is defined by:

$$V = V^{\mathcal{W}} \text{ on } \overline{\mathcal{W}}, \quad V = V^{\mathcal{S}_1} \text{ on } \overline{\mathcal{S}}_1, \quad V = V^{\mathcal{S}_2} \text{ on } \overline{\mathcal{S}}_2. \tag{3.37}$$

The rest of this section is devoted to verifying that V is a classical solution of the HJB variational inequality (3.9) and thus concluding the proof of Theorem 3.2.1 by an application of the martingale optimality principle. We first formalize the heuristic verification from Section 3.2.

3.5.1 Martingale optimality principle

Recall that v is the value function of the optimal liquidation problem (cf. (3.5)).

Proposition 3.5.1 (Martingale optimality principle)**.** *Consider a $C^{2,1}$ function $V : \mathbb{R} \times [0, \infty) \to [0, \infty)$ with the following properties:*

1. *For every $\Theta_{0-} \geq 0$, there exist constants C_1, C_2 so that*

$$V(y, \theta) \leq C_1 \exp(C_2 y) \vee 1 \quad \textit{for all } (y, \theta) \in \mathbb{R} \times [0, \Theta_{0-}];$$

2. *For every $\Theta_{0-} \geq 0$ and $A \in \mathcal{A}(\Theta_{0-})$, the process G from (3.7) is a supermartingale, where $Y = Y^{y,A}$ is defined in (3.3), and additionally $G_0(y; A) \leq G_{0-}(y; A)$.*

Then we have $\overline{S}_0 \cdot V(y,\theta) \geq v(y,\theta)$.
Moreover, if there exists $A^* \in \mathcal{A}(\Theta_{0-})$ *such that* $G(y;A^*)$ *is a martingale and* $G_0(y;A^*) = G_{0-}(y;A^*)$ *holds, then we have* $\overline{S}_0 V(y,\theta) = v(y,\theta)$ *and* $v(y,\theta) = \mathbb{E}[L_\infty(y;A^*)]$ *for* $\Theta_{0-} = \theta \geq 0$. *In this case, any strategy* A *for which* $G(y;A)$ *is not a martingale would be suboptimal.*

Proof. By the supermartingale property we have for every $T \geq 0$

$$\begin{aligned}
\overline{S}_0 V(Y_{0-},\Theta_{0-}) &\geq \mathbb{E}[G_0(y;A)] \geq \mathbb{E}[L_T(y;A) + e^{-\gamma T}\overline{S}_T V(Y_T,\Theta_T)] \\
&= \mathbb{E}[L_T(y;A)] + e^{-\gamma T}\mathbb{E}[\overline{S}_T V(Y_T,\Theta_T)] \\
&= \mathbb{E}[L_T(y;A)] + e^{-\delta T}\overline{S}_0 \mathbb{E}[\mathcal{E}(\sigma W)_T V(Y_T,\Theta_T)].
\end{aligned} \tag{3.38}$$

By monotone convergence, the first summand in (3.38) tends to $\mathbb{E}[L_\infty(y;A)]$ for $T \to \infty$. To see that the second summand converges to 0, consider the Ornstein-Uhlenbeck process $\mathrm{d}X_t = -\beta X_t \,\mathrm{d}t + \hat{\sigma}\,\mathrm{d}B_t,\ X_0 = y$. An application of Itô's formula gives

$$e^{\beta t}(Y_t - X_t) = \int_{[0,t]} e^{\beta u}\,\mathrm{d}\Theta_u \quad \forall t \geq 0. \tag{3.39}$$

Since Θ is non-increasing, we conclude $Y_t \leq X_t$ for all $t \geq 0$. Let $p,q > 1$ be conjugate, i.e. $1 = 1/q + 1/p$. Using Hölder's inequality and the bound on V,

$$\begin{aligned}
\mathbb{E}\big[\mathcal{E}(\sigma W)_T V(Y_T,\Theta_T)\big] &\leq \mathbb{E}\big[\mathcal{E}(\sigma W)_T^p\big]^{1/p}\mathbb{E}\big[V(Y_T,\Theta_T)^q\big]^{1/q} \\
&= \mathbb{E}\big[\exp\big(p\sigma W_T - \tfrac{1}{2}p\sigma^2 T\big)\big]^{1/p}\mathbb{E}[V(Y_T,\Theta_T)^q]^{1/q} \\
&= \mathbb{E}[\mathcal{E}(p\sigma W)_T]^{1/p}\exp\Big(\tfrac{1}{p}\big(\tfrac{1}{2}p^2\sigma^2 T - \tfrac{1}{2}p\sigma^2 T\big)\Big)\mathbb{E}\big[V(Y_T,\Theta_T)^q\big]^{1/q} \\
&= \exp\Big(\frac{p-1}{2}\sigma^2 T\Big)\mathbb{E}[V(Y_T,\Theta_T)^q]^{1/q} \\
&\leq \exp\Big(\frac{p-1}{2}\sigma^2 T\Big)\mathbb{E}[C_1^q \exp(qC_2 Y_T) \vee 1]^{1/q} \\
&\leq \exp\Big(\frac{p-1}{2}\sigma^2 T\Big)\mathbb{E}[C_1^q \exp(qC_2 X_T) \vee 1]^{1/q}.
\end{aligned}$$

Using the fact that X is a Gaussian process with mean $\mathbb{E}[X_T] = ye^{-\beta T}$ and variance $\mathrm{Var}(X_T) = \frac{\hat{\sigma}^2}{2\beta}(1 - e^{-2\beta T})$, we get for $K := \mathbb{E}[C_1^q \exp(qC_2 X_T) \vee 1]$ that

$$\begin{aligned}
K &\leq 1 + C_1^q \exp\Big(qC_2\mathbb{E}[X_T] + \frac{1}{2}q^2 C_2^2 \,\mathrm{Var}(X_T)\Big) \\
&\leq 1 + C_1^q \exp\Big(qC_2 y + \frac{\hat{\sigma}^2}{4\beta}q^2 C_2^2\Big).
\end{aligned}$$

This bound on K is independent of T. Now choosing $p > 1$ such that $\frac{p-1}{2}\sigma^2 < \delta$ ensures that $\exp(-\delta T)\exp(\frac{p-1}{2}\sigma^2 T)$ is exponentially decreasing in T, and thus the second summand in (3.38) converges to 0 for $T \to \infty$. This implies that $\overline{S}_0 V(y,\theta) \geq \mathbb{E}[L_\infty(y;A)]$ for all $A \in \mathcal{A}(\theta)$ and yields the first part of the claim. The second part follows similarly by noting that, if $A^* \in \mathcal{A}(\theta)$ is such that $G(y;A^*)$ is a martingale and $G_0(y;A) = G_{0-}(y;A)$, then we have equalities instead of inequalities in the estimates leading to (3.38). By taking $T \to \infty$ we conclude that $\overline{S}_0 V(y,\theta) = \mathbb{E}[L_\infty(y;A^*)]$. Since $\overline{S}_0 V(y,\theta) \geq v(y,\theta)$ by the first part of the claim, we deduce the optimality of A^*. $\square$

To justify later why the stochastic integrals in (3.8) are true martingales, we need the following technical

Lemma 3.5.2. *Let $\Theta_{0-} \geq 0$ be given and $F \in C^{2,1}(\mathbb{R} \times [0,\infty);\mathbb{R})$ be such that there exist constants $C_1, C_2 \geq 0$ with $|F(y,\theta)| \leq C_1 \exp(C_2 y) \vee 1$ for all $(y,\theta) \in \mathbb{R} \times [0,\Theta_{0-}]$. For an admissible strategy $A \in \mathcal{A}(\Theta_{0-})$ let $Y^A =: Y$ denote the impact process defined by* (3.3) *for $y \in \mathbb{R}$. Then the stochastic integral processes*

$$\int_0^\cdot \overline{S}_u F(Y_u,\Theta_u)\,\mathrm{d}B_u \quad \textit{and} \quad \int_0^\cdot \overline{S}_u F(Y_u,\Theta_u)\,\mathrm{d}W_u \quad \textit{are true martingales.}$$

Proof. It suffices to check $\mathbb{E}[\int_0^t \overline{S}_u^2 \exp(2C_2 Y_u)\,\mathrm{d}u] < \infty$ for every $t \geq 0$ by the exponential growth of F. Consider the Ornstein-Uhlenbeck process X given by $\mathrm{d}X_t = -\beta X_t\,\mathrm{d}t + \hat{\sigma}\,\mathrm{d}B_t$, with $X_0 = y$. As in the proof of Proposition 3.5.1 (see (3.39)), we have $Y_t \leq X_t$ for all $t \geq 0$. In particular,

$$\begin{aligned}
\mathbb{E}\Big[\int_0^t \overline{S}_u^2 \exp(2C_2 Y_u)\,\mathrm{d}u\Big] &\leq \mathbb{E}\Big[\int_0^t \overline{S}_u^2 \exp(2C_2 X_u)\,\mathrm{d}u\Big] \\
= \int_0^t \mathbb{E}[\overline{S}_u^2 \exp(2C_2 X_u)]\,\mathrm{d}u &\leq \int_0^t \sqrt{\mathbb{E}[\overline{S}_u^4]\mathbb{E}[\exp(4C_2 X_u)]}\,\mathrm{d}u < \infty,
\end{aligned}$$

using the Cauchy-Schwarz inequality and the fact that X is a Gaussian process. $\square$

3.5.2 Verification and proof of Theorem 3.2.1

Now we verify that V is a classical solution of the variation inequality (3.9) with the boundary condition $V(y,0) = 0$ for all $y \in \mathbb{R}$. That $V(y,0) = 0$ is clear because $M_1(y_0) = 0$. The rest will be split into several lemmas.

Lemma 3.5.3 (Smooth pasting)**.** *Let $(y_b,\theta_b) \in \overline{\mathcal{W}} \cap \overline{\mathcal{S}}$. Then*

$$\Phi(y_b)C'(\theta_b) + \Phi'(y_b)C(\theta_b) = f(y_b)\,, \tag{3.40}$$

$$\Phi'(y_b)C'(\theta_b) + \Phi''(y_b)C(\theta_b) = f'(y_b)\,. \tag{3.41}$$

Proof. This follows easily from $C(\theta_b) = M_1(y_b)$ and $C'(\theta_b) = M_2(y_b)$, see the definition of C and (3.29), together with the definitions of M_1 and M_2, see (3.28). Note that when $(y_b,\theta_b) = (y_0,0)$ we take the right derivative of C at 0 and the equalities still hold true. $\square$

Remark 3.5.4. It might be interesting to point out that (3.40) and (3.41) are sufficient to derive the boundary between the sell and the wait regions. Indeed, solving (3.40) – (3.41) with respect to $C(\theta_b)$ and $C'(\theta_b)$, it is easy to see that $C(\theta_b) = M_1(y_b)$ and

$C'(\theta_b) = M_2(y_b)$. On the other hand, by the chain rule one gets $\theta'(y_b)C'(\theta_b) = M_1'(y_b)$ and thus we derive for the boundary parametrization $\theta(\cdot) = \mathrm{y}^{-1}(\cdot)$ in the appropriate range

$$\theta'(y_b) = \frac{M_1'}{M_2}(y_b),$$

which gives the ODE for the boundary in (3.29). To get the initial condition y_0, note that the boundary condition $V(\cdot, 0) \equiv 0$ gives $C(0) = 0$, i.e. $M_1(y_0) = 0$, exactly as in Lemma 3.4.3. Thus, one could derive the candidate boundary function $\mathrm{y}(\cdot)$ after assuming sufficient smoothness of the function V along the boundary. This is similar to the classical approach in the singular stochastic control literature, cf. [KS86, Section 6]. The reason why we chose the seemingly longer derivation via calculus of variation techniques is the local (one-sided) optimality that we derived in Theorem 3.4.6 and that will be crucial in our verification of the inequalities of the candidate value function in the sell region, see Lemma 3.5.7. Even in the special case of $\lambda(\cdot)$ being constant, a more direct approach to verify the variational inequality is suggesting new, yet unproven (to our best knowledge), properties for quotients of Hermite functions that might be of independent interest, see Remark 3.5.8.

The smooth-pasting property translates to smoothness of V. Moreover, exponential bound on V and V_y will be needed to rely on the verification results from Section 3.5.1.

Lemma 3.5.5. *The function V is $C^{2,1}(\mathbb{R} \times [0, \infty))$. Moreover, for every Θ_{0-} there exist constants C_1, C_2, that depend on Θ_{0-}, such that both $V(y, \theta)$ and $V_y(y, \theta)$ are non-negative and bounded from above by $C_1 \exp(C_2 y) \vee 1$ for all $(y, \theta) \in \mathbb{R} \times [0, \Theta_{0-}]$.*

Proof. In $\mathcal{W}$, the function V is already $C^{2,1}$ by construction and because $C(\theta) = M_1(\mathrm{y}(\theta))$ is continuously differentiable since $\mathrm{y}(\cdot)$ and $M_1(\cdot)$ are so.

For $(y, \theta) \in \mathcal{S}_1$, set $(y_b, \theta_b) := (y - \Delta(y, \theta), \theta - \Delta(y, \theta))$ and $\Delta := \Delta(y, \theta)$ (recall (3.34)). We have by (3.35) for the first and (3.40) for the second equality

$$\begin{aligned} V_y^{\mathcal{S}_1} &= \Phi'(y_b)C(\theta_b)\,(1 - \Delta_y) + \Phi(y_b)C'(\theta_b)\,(-\Delta_y) + f(y) - f(y_b)\,(1 - \Delta_y) \\ &= \Phi'(y - \Delta)C(\theta - \Delta) + f(y) - f(y - \Delta). \end{aligned} \tag{3.42}$$

Since f, Δ, C and Φ' are continuously differentiable, V_y will also be so. Hence by (3.41),

$$\begin{aligned} V_{yy}^{\mathcal{S}_1} &= \Phi''(y_b)C(\theta_b)\,(1 - \Delta_y) + \Phi'(y_b)C'(\theta_b)\,(-\Delta_y) + f'(y) - f'(y_b)\,(1 - \Delta_y) \\ &= V_{yy}^{\mathcal{W}}(y_b, \theta_b) + f'(y) - f'(y_b), \end{aligned} \tag{3.43}$$

which is continuous. On the other hand, by (3.35) and (3.41) we have

$$\begin{aligned} V_\theta^{\mathcal{S}_1}(y, \theta) &= \Phi'(y_b)C(\theta_b)(-\Delta_\theta) + \Phi(y_b)C'(\theta_b)(1 - \Delta_\theta) - f(y_b)(-\Delta_\theta) \\ &= \Phi(y_b)C'(\theta_b), \end{aligned} \tag{3.44}$$

which is continuous. For $(y,\theta) \in \overline{\mathcal{W}} \cap \overline{\mathcal{S}}$ on the boundary, the left derivative w.r.t. y is

$$\lim_{x \searrow 0} \frac{1}{x}\big(V(y,\theta) - V(y-x,\theta)\big) = \Phi(y)C(\theta),$$

while the right derivative is again given by (3.42) and is equal to the left derivative since $\Delta(y,\theta) = 0$ in this case. Hence, V is continuously differentiable w.r.t. y on the boundary with derivative $V_y(y,\theta) = \Phi'(y)C(\theta)$. Similarly, the left derivative of V_y on the boundary is $\Phi''(y)C(\theta)$ and is equal to the right derivative which is given by (3.43) with $y = y_b$. The left derivative of V w.r.t. θ on the boundary is equal to the right derivative (given by (3.44)). Therefore, V is $C^{2,1}$ inside $\overline{\mathcal{W}} \cup \mathcal{S}_1$.

For $(y,\theta) \in \mathcal{S}_2$, we have that $V_y^{\mathcal{S}_2} = f(y) - f(y-\theta)$, $V_{yy}^{\mathcal{S}_2} = f'(y) - f'(y-\theta)$ and $V_\theta^{\mathcal{S}_2} = f(y-\theta)$ by (3.36), which are all continuous. On the boundary between $\mathcal{S}_1$ and $\mathcal{S}_2$, the left derivative of V w.r.t. y is given by (3.42) while the right derivative is $f(y) - f(y_0)$. Since $\theta - \Delta = 0$ in this case and $C(0) = 0$, they are equal and hence V is continuously differentiable w.r.t. y there; similarly for V_{yy}. The left derivative of V w.r.t. θ there is given by (3.44) with $(y_b, \theta_b) = (y_0, 0)$. The right derivative w.r.t. θ is $f(y-\theta) = f(y_0)$. They are equal by (3.41) and $C(0) = 0$. Therefore, V is $C^{2,1}$ on $\overline{\mathcal{S}}_1 \cup \mathcal{S}_2$. It remains to check smoothness on $\{(y,0) : y \in \mathbb{R}\}$. The derivatives w.r.t. y there are 0. V is continuously differentiable w.r.t. θ in this case because $\mathrm{y}(\cdot)$, C, and Δ are continuously differentiable w.r.t. θ also at $\theta = 0$ (we consider the right derivatives in this case).

To conclude the proof, the bound of V and V_y can be argued as follows. In the wait region, which is contained in $(-\infty, y_0] \times [0,\infty)$, we have $V(y,\theta) = C(\theta)\Phi(y)$ and $V_y(y,\theta) = C(\theta)\Phi'(y)$. Since Φ, Φ' are strictly increasing in y (see (3.6) and [Leb72, Chapter 10] for properties of the Hermite functions), V and V_y will be bounded by a constant there. Now, in the sell region we have $f - V_y - V_\theta = 0$. However, $V_\theta > 0$ because in $\mathcal{S}_1$ (3.44) holds and $C'(\theta_b) = M_2(\mathrm{y}(\theta_b)) > 0$, while in $\mathcal{S}_2$ we have that $V_\theta(y,\theta) = f(y-\theta) > 0$. Similarly, $V_y > 0$ in the sell region. Therefore, $0 < V_y(y,\theta) < f(y) \le \exp(\lambda_\infty y) \vee 1$ by Assumption C4. Hence, integrating in y gives $V(y,\theta) \le V(0,\theta) + \exp(\lambda_\infty y)/\lambda_\infty$ for $y \ge 0$, which implies $V(y,\theta) \le C_1 \exp(C_2 y) \vee 1$ for appropriate constants C_1, C_2. □

Next we prove that V solves the variational inequality (3.9).

Lemma 3.5.6. *The function $V^{\mathcal{W}} : \overline{\mathcal{W}} \to [0,\infty)$ from* (3.33) *satisfies*

$$\mathcal{L}V^{\mathcal{W}}(y,\theta) = 0 \quad \textit{and} \quad f(y) < V_y^{\mathcal{W}}(y,\theta) + V_\theta^{\mathcal{W}}(y,\theta) \textit{ for } y < \mathrm{y}(\theta).$$

Proof. By (3.27), we have $V_\theta^{\mathcal{W}} = \Phi(y)M_1'(\mathrm{y}(\theta))\mathrm{y}'(\theta) = \Phi(y)M_2(\mathrm{y}(\theta))$ and $V_y^{\mathcal{W}} = \Phi'(y)M_1(\mathrm{y}(\theta))$. Recall that at $y = \mathrm{y}(\theta)$ by (3.40) we have $V_y^{\mathcal{W}} + V_\theta^{\mathcal{W}} = f(\mathrm{y}(\theta))$. Now consider $y < \mathrm{y}(\theta)$. By Lemma 3.4.2, we then have $M_1(y) > M_1(\mathrm{y}(\theta))$ giving

$$\Big(\frac{f}{\Phi}\Big)'(y) > \Big(\frac{\Phi'}{\Phi}\Big)'(y)M_1\big(\mathrm{y}(\theta)\big) = \frac{\mathrm{d}}{\mathrm{d}y}\bigg(M_1\big(\mathrm{y}(\theta)\big)\frac{\Phi'(y)}{\Phi(y)} + M_2\big(\mathrm{y}(\theta)\big)\bigg).$$

Therefore, $y \mapsto (f - V_y^{\mathcal{W}}(y,\theta) + V_\theta^{\mathcal{W}}(y,\theta))/\Phi(y)$ is increasing in y. Since at $y = \mathrm{y}(\theta)$ it equals to 0, we get the claimed inequality. □

It remains to verify the inequality in the sell region. The proof is more subtle and that is where Theorem 3.4.6 plays a crucial role. Recall Assumption 3.1.2 and note that y_∞ from Lemma 3.4.3 is unique by condition C3.

Lemma 3.5.7. *The functions $V^{\mathcal{S}_1}$ and $V^{\mathcal{S}_2}$ satisfy on $\overline{\mathcal{S}}_1$ and $\mathcal{S}_2$ respectively*

$$\mathcal{L}V^{\mathcal{S}_1} \leq 0, \quad \mathcal{L}V^{\mathcal{S}_2} < 0.$$

Moreover, the inequality inside $\overline{\mathcal{S}}_1$ is strict except on the boundary between the wait region and the sell region ($\overline{\mathcal{W}} \cap \overline{\mathcal{S}}_1$) where we have equality.

Proof. First consider region $\overline{\mathcal{S}}_1$. Recall from Lemma 3.5.5 (see (3.42) – (3.43)) that in this case

$$\begin{aligned} V_y^{\mathcal{S}_1}(y,\theta) &= V_y^{\mathcal{W}}(y-\Delta,\theta-\Delta) + f(y) - f(y-\Delta), \\ V_{yy}^{\mathcal{S}_1}(y,\theta) &= V_{yy}^{\mathcal{W}}(y_b,\theta_b) + f'(y) - f'(y_b), \end{aligned}$$

where $y = y_b + \Delta(y,\theta)$ and $\theta = \theta_b + \Delta(y,\theta)$. Fix $(y_b,\theta_b) \in \overline{\mathcal{W}} \cap \overline{\mathcal{S}}_1$ and consider the perturbation $\Delta \mapsto (y,\theta) = (y_b+\Delta, \theta_b+\Delta)$. Set

$$\begin{aligned} h(\Delta) &:= \mathcal{L}V^{\mathcal{S}_1}(y_b+\Delta,\theta_b+\Delta) \\ &= \tfrac{\hat\sigma^2}{2} V_{yy}^{\mathcal{W}}(y_b,\theta_b) - \tfrac{\hat\sigma^2}{2} f'(y_b) + \sigma\rho\hat\sigma V_y^{\mathcal{W}}(y_b,\theta_b) - \sigma\rho\hat\sigma f(y_b) - \delta V^{\mathcal{W}}(y_b,\theta_b) \\ &\quad + \tfrac{\hat\sigma^2}{2} f'(y) - \beta y V_y^{\mathcal{W}}(y_b,\theta_b) + \beta y f(y_b) + (\sigma\rho\hat\sigma - \beta y) f(y) - \delta \int_{y_b}^{y} f(x)\,\mathrm{d}x\,. \end{aligned}$$

Note that $h(0) = 0$ by Lemma 3.5.6 and to show $h(\Delta) < 0$ for $\Delta > 0$, it suffices to prove $h'(\Delta) < 0$ for all $\Delta > 0$. We have for all $\Delta \geq 0$ at $y = y_b + \Delta$ that

$$\begin{aligned} h'(\Delta) &= \beta\big(f(y_b) - V_y^{\mathcal{W}}(y_b,\theta_b)\big) \\ &\quad + f(y)\underbrace{\left(\frac{\hat\sigma^2}{2}\frac{f''(y)}{f(y)} - (\beta+\delta) + (\sigma\rho\hat\sigma - \beta y)\frac{f'(y)}{f(y)}\right)}_{=k(y)}, \end{aligned}$$

where at $\Delta = 0$ we consider the right derivative $h'(0+)$. Now we show that $k(y) < 0$ for all $y \geq y_\infty$. To this end, recall that Φ is a solution of the ODE $\delta\Phi(x) = \frac{\hat\sigma^2}{2}\Phi''(x) + (\sigma\rho\hat\sigma - \beta x)\Phi'(x)$. Differentiating w.r.t. x and dividing by $\Phi'(x)$ yields

$$0 = \frac{\hat\sigma^2}{2}\left(\frac{\Phi''(x)}{\Phi'(x)}\right)' + \frac{\hat\sigma^2}{2}\frac{\Phi''(x)^2}{\Phi'(x)^2} - (\beta+\delta) + (\sigma\rho\hat\sigma - \beta x)\frac{\Phi''(x)}{\Phi'(x)}$$

So at the left end y_∞ of our boundary, we have

$$k(y_\infty) = \frac{\hat\sigma^2}{2}\left(\frac{f'}{f}\right)'(y_\infty) + \frac{\hat\sigma^2}{2}\frac{\Phi''(y_\infty)^2}{\Phi'(y_\infty)^2} - (\beta+\delta) + (\sigma\rho\hat\sigma - \beta y_\infty)\frac{\Phi''(y_\infty)}{\Phi'(y_\infty)}$$

$$= \frac{\hat{\sigma}^2}{2}\Big(\frac{f'}{f}\Big)'(y_\infty) - \frac{\hat{\sigma}^2}{2}\Big(\frac{\Phi''}{\Phi'}\Big)'(y_\infty) < 0 \tag{3.45}$$

by Assumption C3. With Assumption C5 we get $k(y) < 0$ for every $y \geq y_\infty$.

In particular, $k(y_b + \Delta) < 0$ for all $\Delta \geq 0$. Since f is positive and increasing, the product $\Delta \mapsto (fk)(y_b + \Delta)$ is decreasing. Therefore, proving $h'(0+) \leq 0$ is sufficient to show the inequality in $\mathcal{S}_1$. To stress the dependence of h on the point $(y_b, \theta_b) = (\mathrm{y}(\theta_b), \theta_b)$, we also write $h(\Delta) = h_{\theta_b}(\Delta)$. Note that $h_\theta(\Delta)$ is continuous in θ and Δ on $[0,\infty) \times [0,\infty)$.

Assume $h'_{\theta_b}(0+) > 0$ at some boundary point (y_b, θ_b) with $\theta_b > 0$. By continuity of h' on θ and Δ there exists some $\varepsilon > 0$ such that $\mathcal{L}V^{\mathcal{S}_1} > 0$ on $U := \overline{\mathcal{S}}_1 \cap B_\varepsilon(y_b, \theta_b)$. This will lead to a contradiction to the fact that the candidate boundary is a (one-sided) strict local maximizer of our stochastic optimization problem with strategies described by the local times of reflected diffusions, see Theorem 3.4.6.

Indeed, fix $\Theta_0 > \theta_b + \varepsilon$ and consider a perturbation $\tilde{\mathrm{y}}(\cdot) \in C^1$ of the boundary $\mathrm{y}(\cdot)$ which satisfies the conditions of Theorem 3.4.6 and $\mathrm{y}(\theta) < \tilde{\mathrm{y}}(\theta) \leq y_0$ in $(\tilde{\mathrm{y}}(\theta), \theta) \in U$ and such that $\tilde{\mathrm{y}}$ and y coincide outside of U. For the corresponding reflection strategies $\tilde{A} := A^{\mathrm{refl}}(\tilde{\mathrm{y}}, \Theta_0)$ and $A := A^{\mathrm{refl}}(\mathrm{y}, \Theta_0)$ denote by $\tilde{\Theta}_t := \Theta_0 - \tilde{A}_t$ and $\Theta_t := \Theta_0 - A_t$ their asset position processes. The liquidation times of $\tilde{A}$ and A are $\tilde{\tau} := \inf\{t \geq 0 : \tilde{A}_t = \Theta_0\}$ and $\tau := \inf\{t \geq 0 : A_t = \Theta_0\}$, respectively. By Theorem 3.3.3 (see also the discussion after (3.11)), we have $T := \tilde{\tau} \vee \tau < \infty$ a.s. Fix initial impact $Y^{\tilde{A}}_{0-} = Y^{A}_{0-} = \mathrm{y}(\Theta_0)$. To compare the strategies A and $\tilde{A}$, consider the processes $G(\mathrm{y}(\Theta_0); A)$ and $G(\mathrm{y}(\Theta_0); \tilde{A})$ from (3.7) for our candidate value function (which is $C^{2,1}$ by Lemma 3.5.5). Since $V(\cdot, 0) = 0$, we have $L_T(\tilde{A}) = G_T(\tilde{A})$ and $L_T(A) = G_T(A)$. However, since $(Y^{\tilde{A}}, \tilde{\Theta})$ spends a positive amount of time in the region $\{\mathcal{L}V > 0\}$ until time T and always remains in the region $\{\mathcal{L}V \geq 0\}$, the perturbed strategy $\tilde{A}$ generates larger proceeds (in expectation) than A.

Indeed, by (3.8) applied for $G(\tilde{A})$ and $G(A)$, using monotone convergence (twice) and arguments as in the proof of Proposition 3.5.1 for the first equality (by (3.19) expected proceeds are bounded), and Lemma 3.5.2 for the stochastic integrals in the second line (noting the growth condition from Lemma 3.5.5), we get

$$\begin{aligned}
\mathbb{E}[L_\infty(\tilde{A}) - L_\infty(A)] &= \lim_{n\to\infty} \mathbb{E}[G_{n\wedge T}(\tilde{A}) - G_{n\wedge T}(A)] \\
&= \lim_{n\to\infty} \mathbb{E}\Big[\int_0^{n\wedge T} \dots \,\mathrm{d}W_t + \int_0^{n\wedge T} \dots \,\mathrm{d}B_t + \int_0^{n\wedge T} \mathcal{L}V(Y^{\tilde{A}}_t, \tilde{\Theta}_t)\,\mathrm{d}t\Big] \\
&= \mathbb{E}\Big[\int_0^{T} \mathcal{L}V(Y^{\tilde{A}}_t, \tilde{\Theta}_t)\,\mathrm{d}t\Big] > 0\,.
\end{aligned}$$

This contradicts Theorem 3.4.6, so $h'(0+) \leq 0$ and hence the inequality in $\mathcal{S}_1$ must hold.

It remains to consider the case $(y, \theta) \in \overline{\mathcal{S}}_2$, where $V^{\mathcal{S}_2}_y = f(y) - f(y - \theta)$ and $V^{\mathcal{S}_2}_{yy} = f'(y) - f'(y - \theta)$. Fix $y - \theta =: a \geq y_0$ and consider $\mathcal{L}V^{\mathcal{S}_2}$ as a function of θ. We have

$$\mathcal{L}V^{S_2}(y,\theta) = \frac{\hat{\sigma}^2}{2}\big(f'(a+\theta) - f'(a)\big) + \big(\sigma\rho\hat{\sigma} - \beta(a+\theta)\big)\big(f(a+\theta) - f(a)\big) \\ - \delta \int_a^{a+\theta} f(x)\,\mathrm{d}x.$$

Differentiating the right-hand side w.r.t. θ we get $f(a+\theta)k(a+\theta)$, which is again decreasing in θ because $a \geq y_0$. Since at $\theta = 0$ we have $\mathcal{L}V^{S_2}(y,\theta) = 0$ we deduce the desired inequality. $\square$

Remark 3.5.8. In the particular case when $\lambda = f'/f$ is constant, a more direct approach based on straightforward calculations leads to a conjecture on a property for quotients of Hermite functions. More precisely, to prove $h'(0+) \leq 0$ in this case it turns out to be sufficient to verify that the map $y_b \mapsto h'(0)$ is monotone in $[y_\infty, y_0]$, because at y_∞ and y_0 one can check that $h'(0+) < 0$. The monotonicity in y_b would then follow from the following conjectured property of the Hermite functions:

$$\text{For every } \nu < 0, \text{ the function} \quad x \mapsto \frac{(H_{\nu-1}(x))^2}{H_\nu(x)H_{\nu-2}(x)} \quad \text{is decreasing.}$$

Numerical computations indicate the validity of the this property but, to our best knowledge, it is not yet proven and may be of independent interest. Note that such quotients of special functions are related to so called Turan-type inequalities, cf. [BI13].

Now we have all the ingredients in place to complete the

Proof of Theorem 3.2.1. The function V constructed in (3.37) is a classical solution of the variational inequality (3.9) because of Lemmas 3.5.5, 3.5.6 and 3.5.7. Thus, for each admissible strategy A the process $G(y;A)$ from (3.7) is a supermartingale with $G_0(y;A) \leq G_{0-}(y;A)$: the growth condition on V_y and V from Lemma 3.5.5 guarantees that the stochastic integral processes in (3.8) are true martingales by an application of Lemma 3.5.2, while the variational inequality gives the supermartingale property on $[0-,\infty)$. Moreover, for the described strategy A^*, whose existence and uniqueness on $[\![0,\tau]\!]$ follows from classical results, cf. Remark 3.3.2, the process $G(y;A^*)$ is a true martingale with $G_0(y;A^*) = G_{0-}(y;A^*)$ by our construction of V and the validity of the variational inequality in the respective regions. Therefore A^* is an optimal strategy by Proposition 3.5.1. Any other strategy will be suboptimal because the respective inequalities are strict in the sell and wait region, i.e., for any other strategy the process G will be a strict supermartingale.

The Laplace transform formula (3.11) was derived in Theorem 3.3.3 for a $\mathbb{y}$-reflected strategy when the state process starts on the boundary. If the state process starts in $Y_0 = x$ in the wait region, the behavior of the process until time $H^{x\to z}$ when it hits the boundary for the first time (at $z := \mathbb{y}(\Theta_0)$) is independent from future excursions from the boundary, and hence the multiplicative factor in (3.11), see e.g. [RW87, Prop. V.50.3]: for $x < z \in \mathbb{R}$ and $\alpha > 0$, $\mathbb{E}[\exp(-\alpha H^{x\to z})] = \Phi_\alpha(x)/\Phi_\alpha(z)$. $\square$

4 Superhedging with transient impact of non-covered and covered options

In this chapter, we solve the superhedging problem of European contingent claims by a large trader whose dynamic hedges have a transient (and possibly as well purely permanent) impact on the prices of the underlying asset. We consider the multiplicative price impact specification from Chapter 2 with a fundamental Black-Scholes price for the underlying, to ensure positivity, see Section 4.1 for details.

In Section 4.2 we specify the case of non-covered options, that is when impact from the initial and terminal trades are considered. For such options, in Section 4.3 we formulate the problem of superhedging as a stochastic target problem. We apply stochastic target techniques and geometric dynamic programming in suitably chosen effective coordinates, being related to instantaneous liquidation wealth, to derive in Section 4.4 non-linear pricing pdes whose viscosity solutions characterize the minimal superhedging prices, cf. our main results Theorems 4.4.5 and 4.4.9. In Section 4.5 we explain how our analysis applies directly to a model extension that has also permanent impact. We close our study on pricing and hedging of non-covered options with a numerical illustration in Section 4.6.

In Section 4.7 we consider the case of covered options, that is when initial and terminal impact could be disregarded. It turns out that the pricing pde is of completely different nature, being degenerate with gamma constraints. At the end in Section 4.8 we collect technical proofs delegated from Section 4.4.

4.1 Transient price impact model

This section describes the multiplicative market impact model that we will consider in this chapter. An extension with additional permanent impact is postponed to Section 4.5. We fix a complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with countably generated $\mathcal{F}$, a filtration $\mathbb{F} = (\mathcal{F}_t)_{t\geq 0}$ satisfying the usual conditions, and a Brownian motion W on this filtered probability space. For instance, we could take Ω to be the canonical space of continuous functions on $[0, \infty)$, $\mathbb{P}$ the Wiener measure, W the canonical process, and $\mathbb{F}$ its augmented raw filtration, possibly extended by a sequence of random measures, with $\mathcal{F} = \bigvee_{t\geq 0} \mathcal{F}_t$, see [ST02, Section 2.5].

In the absence of the large trader, the *unaffected* price process $\overline{S}$ of the single risky asset evolves according to the stochastic differential equation

$$\mathrm{d}\overline{S}_t = \overline{S}_t(\mu_t \,\mathrm{d}t + \sigma \,\mathrm{d}W_t), \quad \overline{S}_0 \in \mathbb{R}_+, \tag{4.1}$$

with constant $\sigma > 0$ and bounded progressive process μ. Let the càdlàg adapted process Θ denote the evolution of his holdings in the risky asset. We define the market impact process $Y = Y^\Theta$ pathwise, on the Skorohod space of càdlàg paths, via

$$\mathrm{d}Y_t^\Theta = -h(Y_t^\Theta)\,\mathrm{d}t + \mathrm{d}\Theta_t, \quad Y_{0-} = y \in \mathbb{R}, \tag{4.2}$$

for $h : \mathbb{R} \to \mathbb{R}$ being a Lipschitz continuous function with $\operatorname{sgn}(x)h(x) \geq 0$. When the large trader follows strategy Θ, the risky asset price observed on the market, being the marginal price at which additional infinitesimal quantities could be traded, is

$$S_t^\Theta = S_t = f(Y_t^\Theta)\overline{S}_t, \quad t \geq 0, \tag{4.3}$$

where the price impact function $f : \mathbb{R} \to \mathbb{R}_+$ is increasing and in C^1 with $f(0) = 1$. In particular, $\lambda := f'/f$ is a non-negative and locally integrable C^0 function, satisfying

$$f(x) = \exp\left(\int_0^x \lambda(u)\,\mathrm{d}u\right), \quad x \in \mathbb{R}. \tag{4.4}$$

By the monotonicity of f, the price impact from his trades is adverse to the large trader. During periods where the large trader is inactive, the impact process Y recovers towards its neutral state 0, so that the relative price impact $S/\overline{S} = f(Y)$ w.r.t. the unaffected (fundamental) price $\overline{S}$ is persistent but lessens over time, rendering the impact as *transient*.

Next, we specify the large trader's proceeds (negative expenses) L, which are the variations of his cash account to finance the dynamic holdings Θ in the risky asset. For simplicity, we assume zero interest rates and a riskless asset with constant price 1 as cash, i.e. prices are discounted in units of this numeraire asset. For continuous strategies Θ of finite variation,

$$L(\Theta) = -\int_0^\cdot S^\Theta\,\mathrm{d}\Theta \tag{4.5}$$

are the proceeds. And there is a unique continuous extension of the proceeds functional in (4.5) to general (bounded)[1] semimartingale strategies Θ, that is given by

$$L(\Theta) := \int_0^\cdot F(Y_t^\Theta)\,\mathrm{d}\overline{S}_t - \int_0^\cdot \overline{S}_t(fh)(Y_t^\Theta)\,\mathrm{d}t - (\overline{S}F(Y^\Theta) - \overline{S}_0F(Y_{0-}^\Theta)), \tag{4.6}$$

as shown in Chapter 2 (cf. Theorem 2.2.7), with antiderivative

$$F(x) := \int_0^x f(u)\,\mathrm{d}u, \quad x \in \mathbb{R}. \tag{4.7}$$

More precisely, every (càdlàg) semimartingale can be approximated (in probability) in

[1] Results in Chapter 2 are stated in a more general setup where $\overline{S}$ can have jumps and trading strategies do not need to be bounded and semimartingales. Yet, for the analysis in the current chapter we can restrict to bounded semimartingale strategies.

the Skorokhod's M_1 topology by a sequence of continuous processes of finite variation, see Section 2.2.1 for details, and if semimartingales $\Theta^n \xrightarrow{\mathbb{P}} \Theta$ in $(D([0,T]), M_1)$ for a semimartingale Θ, then $L(\Theta^n) \xrightarrow{\mathbb{P}} L(\Theta)$ in $(D([0,T]), M_1)$. Thus, it is natural to define L by (4.6) as the continuous extension of L from (4.5) to all semimartingales.

Thus, the proceeds from a block trade of size $\Delta\Theta_t$ at time t are given by

$$-\overline{S}_t \int_0^{\Delta\Theta_t} f(Y_{t-}^{\Theta} + x)\,\mathrm{d}x,$$

showing that the price per share that the large trader pays/obtains for a block buy/sell order is between the price before the trade $f(Y_{t-}^{\Theta})\overline{S}_t$ and the price after the trade $f(Y_t^{\Theta})\overline{S}_t$. The form proceeds and price impact from block trades can be interpreted from the perspective of a latent limit order book, where a block trade is executed against available orders in the order book for prices between $f(Y_{t-}^{\Theta})\overline{S}_t$ and $f(Y_{t-}^{\Theta} + \Delta\Theta_t)\overline{S}_t$, see Section 2.4.1. In this sense, Y is a volume effect process in spirit of [PSS11]).

For a self-financing portfolio (β, Θ), in which the dynamic holdings in cash (the riskless asset) and in stock (say, the risky asset) evolve as β and Θ, respectively, the self-financing condition is

$$\beta = \beta_{0-} + L(\Theta).$$

In order to define the wealth dynamics induced by the large trader's strategy, one needs to specify the dynamics of the value of the risky asset position in the portfolio. If the large trader were forced to liquidate his stock position immediately by a single block trade, the instantaneous *liquidation wealth* V_t^{liq} is

$$V_t^{\mathrm{liq}} = V_t^{\mathrm{liq}}(\Theta) := \beta_t + \overline{S}_t \int_0^{\Theta_t} f(Y_t^{\Theta} - x)\,\mathrm{d}x. \tag{4.8}$$

The dynamics for this notion of wealth is mathematically tractable and continuous, satisfying

$$\mathrm{d}V_t^{\mathrm{liq}} = (F(Y_{t-}) - F(Y_{t-} - \Theta_{t-}))\,\mathrm{d}\overline{S}_t - \overline{S}_t(f(Y_{t-}) - f(Y_{t-} - \Theta_{t-}))h(Y_t)\,\mathrm{d}t. \tag{4.9}$$

One obtains from (4.9) absence of arbitrage within the following set of *admissible strategies*

$$\begin{aligned}\mathcal{A}^{\mathrm{NA}} := \big\{(\Theta_t)_{t\geq 0} \mid &\text{ bounded semimartingale, with } \Theta_{0-} = 0\\ &\text{and } \Theta_t = 0 \text{ on } t \in [T,\infty) \text{ for some } T < \infty\big\}.\end{aligned}$$

Proposition 4.1.1. *The market is free of arbitrage up to any finite time horizon $T \in [0,\infty)$ in the sense that there exists no $\Theta \in \mathcal{A}^{NA}$ with $\Theta_t = 0$ on $t \in [T,\infty)$ such that for the self-financing strategy (β,Θ) with $\beta_{0-} = 0$ we have $\mathbb{P}[V_T^{liq} \geq 0] = 1$ and $\mathbb{P}[V_T^{liq} > 0] > 0$.*

Proof. The claim is proven as in Section 2.3. We note that there it was additionally required for admissible strategies that V^{liq} is bounded from below. The latter condition however can be omitted in the current setup of bounded controls. To see this, observe that for any $\Theta \in \mathcal{A}^{\text{NA}}$ there exists an equivalent martingale measure $\mathbb{Q}^{\Theta} \approx \mathbb{P}$ (on $\mathcal{F}_T$), constructed as in [BBF17b, proof of Thm. 4.3], under which the process V^{liq} is a true martingale. □

Unlike to the friction-less situation, there is more than one sensible way to define wealth in an illiquid market with price impact. For instance, for the analysis in Section 4.7 we shall also make use of another notion of *book wealth*. For a strategy with dynamic holdings Θ and β in the risky and the riskless asset, the *book wealth process* is given by

$$V^{\text{book}} := \beta + \Theta S, \tag{4.10}$$

with the risky asset being evaluated at the current (marginal) market prices S currently observed. In illiquid markets, the liquidation wealth equation (4.8) which is achievable by the large trader if he were to unwind his risky asset holdings immediately is usually different from the book wealth.

4.2 Hedging of non-covered options in illiquid markets

We solve in Sections 4.2-4.6 the problem of dynamic hedging for non-covered options, where the issuer who wants to hedge the option is to receive the option premium in cash, whereas Section 4.7 discusses the related problem for the (less familiar) case of covered options. For the latter, a part of the premium (the initial 'delta') is to be paid in shares of the underlying risky asset by the buyer, at the discretion of the issuer, and at maturity the buyer accepts a mixture of the riskless and risky asset, evaluated at the current market prices, as a payout.

In illiquid markets, it is relevant to distinguish between *cash settlement* and *physical settlement* of the option payoff because, in contrast to frictionless models with infinite liquidity, moving funds by performing lump trade between the bank account and the risky asset account induces additional costs due to illiquidity, and can change the price of the underlying and thereby affect the option's payoff. We consider European options, or contingent claims, with fixed maturity $T \geq 0$.

Definition 4.2.1. *European option with maturity T is specified by a measurable map*

$$(g_0, g_1) : (s, y) \in \mathbb{R}_+ \times \mathbb{R} \mapsto (g_0(s, y), g_1(s, y)) \in \mathbb{R} \times \mathbb{R}$$

representing the payoff, where g_0 is the cash-settlement part and g_1 is the physical-delivery part at maturity. It entitles its holder the payment of $g_0(S_T, Y_T)$ in cash and $g_1(S_T, Y_T)$ in risky asset, where (S_T, Y_T) is the risky asset price and the level of market impact at maturity.

In the remainder, $T \geq 0$ denotes the (fixed) maturity when it comes to European options. The seller of an non-covered European option with payoff (g_0, g_1) needs to hedge against possible losses from his obligation to deliver the payoff at maturity. Among his admissible trading strategies Γ (that we will specify more precisely later in Section 4.3.1), he will consider the following trading strategies.

Definition 4.2.2 (Hedging of non-covered option)**.** *A superhedging strategy is a self-financing strategy (β, Θ) with $\Theta \in \Gamma$, $\Theta_{0-} = 0$, and*

$$\beta_T \geq g_0(S_T, Y_T) \quad \text{and} \quad \Theta_T = g_1(S_T, Y_T).$$

Let us point out that a hedging strategy has to deliver exactly the physical component $g_1(S_T, Y_T)$ at maturity, and that any further (long or short) position in the underlying has to be unwound before the options are settled at the resulting price S_T and impact level Y_T. This constrains the hedger's opportunities to manipulate the payoff at maturity by performing block trades (to be unwound shortly after that) that might move the prices in favorable direction, and is different from [BB04]. The (minimal) superhedging price of a non-covered option with payoff (g_0, g_1), which we will denote by $p_{(g_0,g_1)}$, is the minimal (infimum of) initial capital β_{0-} for which such a superhedging strategy (β, Θ) exists.

Options with pure cash settlement are characterized by $g_1 = 0$. In fact, every (reasonable) option can be represented by a payoff with pure cash settlement. Indeed, if Γ is stable under adding an additional jump at terminal time, meaning that $\Theta \in \Gamma$ implies that $\Theta + \Delta \mathbb{1}_{\{T\}} \in \Gamma$ for every $\mathcal{F}_T$-measurable Δ, then every European option can be represented by an option with pure cash settlement. Indeed, for a European option with payoff (g_0, g_1), let for $(s, y) \in \mathbb{R}_+ \times \mathbb{R}$

$$H(s, y) := \inf \left\{ g_0\big(s \tfrac{f(y+\theta)}{f(y)}, y + \theta\big) + s \tfrac{F(y+\theta) - F(y)}{f(y)} \mid \theta = g_1\big(s \tfrac{f(y+\theta)}{f(y)}, y + \theta\big) \right\}. \quad (4.11)$$

The value $H(s, y)$ is the minimal initial capital (in the riskless asset) needed to hedge the payoff (g_0, g_1) with a single (instant) block trade, when before that trade the level of impact is y and there are no holdings in the risky asset whose price is s. Indeed, a block trade of size θ will result in the new price $\tilde{s} = sf(y + \theta)/f(y)$ and impact $\tilde{y} = y + \theta$, it will incur the cost $s(F(y + \theta) - F(y))/f(y)$ and thus will hedge the claim (g_0, g_1) if $\theta = g_1(\tilde{s}, \tilde{y})$ and we have enough capital to pay for the block trade and to cover the cash-delivery part that after the block trade equals $g_0\big(\tilde{s}, \tilde{y}\big)$, see Definition 4.2.2. We have the following result.

Lemma 4.2.3. *For a European option with payoff (g_0, g_1), let H from* (4.11) *be finite and measurable. Then $p_{(g_0,g_1)} = p_{(H,0)}$.*
In the case of λ being constant, if g_0 and g_1 do not depend on y, then H also does not depend on y.

Proof. Suppose that (β, Θ) is a superhedging strategy for (g_0, g_1). This means that $\Theta_T = g_1(S_T, Y_T)$ and $\beta_{0-} + L_T(\Theta) \geq g_0(S_T, Y_T)$. Consider the strategy $\widetilde{\Theta} := \Theta - \Theta_T \mathbb{1}_{\{T\}}$. The price and impact after implementing $\widetilde{\Theta}$ will be $\widetilde{S}_T = S_T f(Y_T - \Theta_T)/f(Y)$ and $\widetilde{Y}_T = Y_T - \Theta_T$ respectively, and the generated proceeds, which in this case will also be equal to $V_T^{\text{liq}}(\widetilde{\Theta})$, are

$$V_T^{\text{liq}}(\widetilde{\Theta}) = \beta_{0-} + L_T(\Theta) + \overline{S}_T(F(Y_T) - F(Y_T - \Theta_T))$$

Hence,

$$\begin{aligned} V_T^{\text{liq}}(\widetilde{\Theta}) &\geq g_0(S_T, Y_T) + \overline{S}_T(F(Y_T) - F(Y_T - \Theta_T)) \\ &= g_0\big(\widetilde{S}_T \tfrac{f(\widetilde{Y}_T + \Theta_T)}{f(\widetilde{Y}_T)}, \widetilde{Y}_T + \Theta_T\big) + \widetilde{S}_T \tfrac{F(\widetilde{Y}_T + \Theta_T) - F(\widetilde{Y}_T)}{f(\widetilde{Y}_T)} \\ &\geq H(\widetilde{S}_T, \widetilde{Y}_T). \end{aligned}$$

Therefore, the self-financing trading strategy $\widetilde{\Theta}$ with initial capital β_{0-} is a superhedging strategy for the European claim with payoff $(H, 0)$, giving that $p_{(g_0,g_1)} \geq p_{(H,0)}$.

To show the reverse inequality, let (β, Θ) be a superhedging strategy for $(H, 0)$, meaning that $\Theta_T = 0$ and $V_T^{\text{liq}}(\Theta) \geq H(S_T, Y_T)$. A measurable selection argument yields that for any $\varepsilon > 0$, there exists an $\mathcal{F}_T$-measurable random variable Θ_T^ε such that $\Theta_T^\varepsilon = g_1\big(s \frac{f(Y_T + \Theta_T^\varepsilon)}{f(Y_T)}, Y_T + \Theta_T^\varepsilon\big)$ and

$$H(S_T, Y_T) + \varepsilon \geq g_0\big(S_T \tfrac{f(Y_T + \Theta_T^\varepsilon)}{f(Y_T)}, Y_T + \Theta_T^\varepsilon\big) + S_T \tfrac{F(Y_T + \Theta_T^\varepsilon) - F(Y_T)}{f(Y_T)}.$$

Thus, the strategy $\widetilde{\Theta}^\varepsilon := \Theta + \Theta_T^\varepsilon \mathbb{1}_{\{T\}}$ with initial capital $\beta_{0-} + \varepsilon$ is superhedging for the claim with payoff (g_0, g_1); indeed, the proceeds generated from $\widetilde{\Theta}^\varepsilon$ are

$$V_T^{\text{liq}}(\Theta) + \varepsilon - S_T \tfrac{1}{f(Y_T)} \left(F(Y_T + \Theta_T^\varepsilon) - F(Y_T)\right)$$

where the last term is the cost of acquiring Θ_T^ε assets. Hence, with reference to the preceding inequality, the generated proceeds from $\widetilde{\Theta}^\varepsilon$ are sufficient to deliver the cash part of the payoff, and also physical part by choice of θ^ε. Since $\varepsilon > 0$ was arbitrary, we conclude $p_{(H,0)} \geq p_{(g_0,g_1)}$, and thus the claim.

If λ is constant, then we have $f(x) = \exp(\lambda x)$ and thus $(F(y + \theta) - F(y))/f(y) = F(\theta)$ and $f(y + \theta)/f(y) = f(\theta)$. In particular, H does not depend on y, if g_0 and g_1 do not. □

Example 4.2.4. 1. A cash-settled European call option with strike K is specified by the payoff $(g_0(s, y), g_1(s, y)) = ((s - K)^+, 0)$.

2. In comparison, a European call option with strike K and physical settlement has the payoff $(-K\mathbb{1}_{\{s \geq K\}}, \mathbb{1}_{\{s \geq K\}})$. Note that although the payoff profile (g_0, g_1) does not depend on the level of impact y, the equivalent pure cash settlement profile H from Lemma 4.2.3 can still depend on y if λ is not constant. Indeed, in general the effect on the relative price change $f(y + \theta)/f(y)$ from a block trade θ depends on the level y of impact before the trade (unless $f(x) = \exp(\lambda x)$).

Remark 4.2.5. We discuss an example to demonstrate how the hedging problem for the large trader could be related to hedging in a liquid market but with portfolio constraints, if F from (4.7) is not surjective onto $\mathbb{R}$. In particular, in this case our market model will not be complete in the sense that not every contingent claim can be perfectly replicated. A prototypical example is the special case of pure permanent impact, i.e. $h \equiv 0$, constant λ, i.e. $f(x) = \exp(\lambda x)$, and a claim with payoff $(H, 0)$, i.e. only cash settlement. Hence, we are in the setup of [BB04] with the smooth family of semimartingales $P(x,t) := \exp(\lambda x)\overline{S}_t$. If $Y_{0-} = 0$ and $\lambda = 1$, (4.9) reduces to

$$\mathrm{d}V_t^{\mathrm{liq}} = (\exp(\Theta_t) - 1)\,\mathrm{d}\overline{S}_t.$$

Note that by our assumption on the hedging strategies in Definition 4.2.2, any hedging Θ will satisfy $\Theta_T = 0$, and hence at maturity $S_T = \overline{S}_T$ and $Y_T = Y_{0-} = 0$. Thus, the superreplication condition becomes $V_T^{\mathrm{liq}}(\Theta) \geq H(\overline{S}_T, 0)$. This means that, after a reparametrization $\Theta \mapsto \exp(\Theta) - 1$ of the strategies, the pricing problem in this large investor model is equivalent to the pricing problem in the respective frictionless model with price process $\overline{S}$ (i.e. for a small investor) and with constraints on *the delta* (greater than -1), i.e. the number of risky assets that a hedging strategy might have. In particular, one should expect that in such situations (when F is not surjective) the pricing equation should contain gradient constraints. Note that this is different from [BB04] because for this particular f their crucial Assumption 5 is violated, and also different from [BLZ16] because their assumption (H2) would not hold in this case.

In the presence of resilience in the market impact ($h \not\equiv 0$), the situation is more delicate since the evolution of the price and impact processes depend on the full history of the trading strategy and thus such a simplification there is not immediate. We will see later in Section 4.4.2 that in the case $f = \exp(\lambda\cdot)$ a lower bound on the delta will also emerge naturally in order to make sense of the pricing equation.

4.3 Superhedging by geometric dynamic programming

In this section, we restate the superhedging problem for non-covered options as a stochastic target problem and are going to show that a Dynamic Programming Principle (DPP for short) holds and the value function could be characterized as the viscosity solution of a suitable pde. More precisely, in our setup a special form of the DPP holds along suitably modified (reduced) state process, that is the *effective* state process that would result from an instant liquidation of the risky position. This change of coordinates will be the main ingredient to derive the pricing pde in Section 4.4.

4.3.1 Stochastic target formulation

We consider strategies that take values in the constraint set $\mathcal{K} \subseteq \mathbb{R}$, for one of the two cases

$$\mathcal{K} = [-K, +\infty) \text{ for some } K > 0, \text{ or} \tag{4.12}$$

$$\mathcal{K} = \mathbb{R}. \tag{4.13}$$

The short-selling constraints (4.12) will be needed when F is not surjective onto $\mathbb{R}$, see Remark 4.2.5, in which case we will consider in Section 4.4.2 $f(x) = \exp(\lambda x)$ for some $\lambda > 0$, while $\mathcal{K} = \mathbb{R}$ will be in force when f is bounded away from 0 and $+\infty$, meaning that the (relative) change of the price from a block trade cannot be arbitrarily big.

Our admissible trading strategies require jumps in order to obtain a DPP; In fact, even without price impact hedging strategies in the Black-Scholes model require initial and terminal block trades. For $k \in \mathbb{N}$, let $\mathcal{U}_k$ denote the set of random $\{0, \dots, k\}$-valued measures ν supported on $[-k, k] \times [0, T]$ that are adapted in the following sense: for every $A \in \mathcal{B}([-k, k])$, the process $t \mapsto \nu(A, [0, t])$ is adapted to the underlying filtration. Note that the elements of $\mathcal{U}_k$ have the representation

$$\nu(A, [0, t]) = \sum_{i=0}^{k} \mathbb{1}_{\{(\delta_i, \tau_i) \in A \times [0,t]\}},$$

where $0 \le \tau_1 < \cdots < \tau_k \le T$ are stopping times and δ_i is a $[-k, k]$-valued $\mathcal{F}_{\tau_i}$-random variable (might take values 0 as well). Consider also $\mathcal{U} := \bigcup_{k \ge 1} \mathcal{U}_k$.

The admissible trading strategies Θ that we will consider are bounded, take values in $\mathcal{K}$ and have the representation

$$\Theta_t = \Theta_{0-} + \int_0^t a_s \, \mathrm{d}s + \int_0^t b_s \, \mathrm{d}W_s + \int_0^t \int_{\mathbb{R}} \delta \nu(\mathrm{d}\delta, \, \mathrm{d}s), \tag{4.14}$$

in which $\Theta_{0-} \in \mathcal{K}$, $\nu \in \mathcal{U}$ and $(a, b) \in \mathcal{A} := \bigcup_{k \ge 1} \mathcal{A}_k$, where for $k \ge 1$

$$\mathcal{A}_k := \{(a, b) \mid a \text{ and } b \text{ are predict. with } |a| \vee |b| \le k, \ \mathrm{d}t \otimes \mathrm{d}\mathbb{P}\text{-a.e.}\}.$$

In this sense, we identify the trading strategies by triplets $(a, b, \nu) \in \mathcal{A} \times \mathcal{U}$. For $k \in \mathbb{N}$ set

$$\Gamma_k := \{(a, b, \nu) \in \mathcal{A}_k \times \mathcal{U}_k \ : \ \Theta \text{ from (4.14) takes values in } \mathcal{K} \cap [-k, k]\}$$

and let $\Gamma := \bigcup_{k \ge 1} \Gamma_k$.

To reformulate the superhedging problem in our market impact model as a stochastic target problem, consider for $(t, z) = (t, s, y, \theta, v) \in [0, T] \times \mathbb{R}_+ \times \mathbb{R} \times \mathcal{K} \times \mathbb{R}$ and $\gamma \in \Gamma$ the (dynamic version of) *the state process*

$$(Z_u^{t,z,\gamma})_{u \in [t,T]} = (S_u^{t,z,\gamma}, Y_u^{t,z,\gamma}, \Theta_u^{t,z,\gamma}, V_u^{\mathrm{liq},t,z,\gamma})_{u \in [t,T]}, \tag{4.15}$$

where the processes $S^{t,z,\gamma}, Y^{t,z,\gamma}, \Theta^{t,z,\gamma}$ and $V^{\mathrm{liq},t,z,\gamma}$ correspond to the price, impact, risky asset position and instantaneous liquidation wealth processes on $[t, T]$ for the control $\Theta^{t,z,\gamma}$ associated with γ (from the decomposition (4.14) on $[t, T]$ instead), when started at time $t-$ at s, y, θ and v respectively.

Following the discussion in Section 4.2, for a non-covered European option with payoff function given by a measurable map $(s, y) \in \mathbb{R}_+ \times \mathbb{R} \mapsto (g_0(s, y), g_1(s, y))$, for a superhedging strategy $\gamma \in \Gamma$ the state process at time T is (a.s.) in the set

$$\mathfrak{G} := \big\{(s, y, \theta, v) \in \mathbb{R}_+ \times \mathbb{R} \times \mathcal{K} \times \mathbb{R} \ : \ \theta = g_1(s, y),\ v - s\big(F(y) - F(y-\theta)\big)/f(y) \geq g_0(s, y)\big\}$$

that we call the *target set*. The superhedging strategies for initial position θ in the risky asset are

$$\mathcal{G}(t, s, y, \theta, v) := \bigcup_{k \geq 1} \mathcal{G}_k(t, s, y, \theta, v)$$

with

$$\mathcal{G}_k(t, s, y, \theta, v) := \{\gamma \in \Gamma_k \ : \ Z_T^{t,s,y,\theta,v,\gamma} \in \mathfrak{G}\}.$$

Following Definition 4.2.2, superhedging strategies will have no risky assets at the beginning and thus the (minimal) superhedging price is

$$w(t, s, y) := \inf_{k \geq 1} w_k(t, s, y), \quad \text{where} \quad w_k(t, s, y) := \inf\{v \ : \ \mathcal{G}_k(t, s, y, 0, v) \neq \emptyset\}. \tag{4.16}$$

Let us point out that the value function depends on the constraint set $\mathcal{K}$ (via the target set $\mathfrak{G}$). Note also that the set of admissible superhedging strategies (identified with $\mathcal{G}(t, s, y, 0, v)$) is a subset of $\mathcal{A}^{\mathrm{NA}}$, meaning that the minimal superhedging price of a positive payoff H, the pure-cash delivery equivalent of (g_0, g_1) as in Lemma 4.2.3, is strictly positive.

4.3.2 Effective coordinates and dynamic programming principle

For stochastic target problems usually a form of the Dynamic Programming Principle (DPP) holds and plays a crucial role in deriving a pde that characterizes the value function (in a viscosity sense). The aim of this section is namely providing a suitable DPP.

First note that the superhedging problem in this form is not time-consistent because in the definition of the minimal superhedging price w, see (4.16), it is assumed that the initial position in the risky asset is 0, while at later times it typically will not be. To have a time-consistent setup, one possible approach could be to make the risky asset position a new variable, i.e. to work with the function $\bar{w}$ defined on $[0, T] \times \mathbb{R}_+ \times \mathbb{R} \times \mathcal{K}$ by

$$\bar{w}(t, s, y, \theta) := \inf_{k \geq 1} \bar{w}_k(t, s, y, \theta) \text{ with } \bar{w}_k(t, s, y, \theta) := \inf\{v \ : \ \mathcal{G}_k(t, s, y, \theta, v) \neq \emptyset\}. \tag{4.17}$$

It turns out that it is possible to reduce the state space by considering the problem in suitable reduced coordinates. In fact, in these new coordinates a DPP holds for the function w instead. We adapt the ideas from [BLZ16] (that seem to appear even earlier in [LL07]) to our setup as follows.

To derive dynamic programming principle for w, we want to compare it at different points in time with the wealth process. Since w assumes zero initial risky assets, it is natural to consider the (fictitious) state process that would prevail if the trader would be forced to liquidate his position in the risky asset immediately (with a block trade). To this end, let

$$\begin{aligned} \mathcal{S}(S_t, Y_t^\Theta, \Theta_t) &:= \overline{S}_t f(Y_t^\Theta - \Theta_t) \qquad \big(= S_t f(Y_t^\Theta - \Theta_t)/f(Y_t^\Theta)\big)\,, \\ \mathcal{Y}(Y_t^\Theta, \Theta_t) &:= Y_t^\Theta - \Theta_t \end{aligned}$$

These processes can be interpreted as follows: $\mathcal{S}(s, y, \theta)$ is the price of the asset that would prevail after θ assets were liquidated, when s and y are the price of the risky asset and the market impact just before the trade, while $\mathcal{Y}(y, \theta)$ would be the state of the impact after this trade. In this sense, for a self-financing trading strategy Θ we refer to the processes $\mathcal{S}(S_t, Y_t^\Theta, \Theta_t)$ and $\mathcal{Y}(Y_t^\Theta, \Theta_t)$ as the *effective price and impact processes*, respectively. Observe that these processes are continuous, although Θ might have jumps.

For the subsequent dynamic programming principle (DPP), see Theorem 4.3.1 below, we will be comparing the instantaneous liquidation wealth V^{liq} with the value function w along the evolution of $(\mathcal{S}(S, Y^\Theta, \Theta), \mathcal{Y}(Y^\Theta, \Theta))$. While the proof for this DPP is mainly following ideas due to [BLZ16, Prop.3.3], we would like to point out that the arguments simplify in technical terms and appear more transparent when expressed in terms of our choice for V^{liq}, instead of V^{book}.

Theorem 4.3.1 (Geometric DPP)**.** *Fix $(t, s, y, v) \in [0, T] \times \mathbb{R}_+ \times \mathbb{R} \times \mathbb{R}$.*

Part 1. *If $v > w(t, s, y)$, then there exists $\gamma \in \Gamma$ and $\theta \in \mathcal{K}$ such that*

$$V_\tau^{liq,t,z,\gamma} \geq w(\tau, \mathcal{S}(S_\tau^{t,z,\gamma}, Y_\tau^{t,z,\gamma}, \Theta_\tau^{t,z,\gamma}), Y_\tau^{t,z,\gamma} - \Theta_\tau^{t,z,\gamma})$$

for all stopping times $\tau \geq t$, where $z = (\mathcal{S}(s, y, -\theta), y + \theta, \theta, v)$.

Part 2. *Let $k \geq 1$. If $v < w_{2k+2}(t, s, y)$, then for every $\gamma \in \Gamma_k$, $\theta \in \mathcal{K} \cap [-k, k]$ and a stopping time $\tau \geq t$ we have*

$$\mathbb{P}\left[\ V_\tau^{liq,t,z,\gamma} > w_k(\tau, \mathcal{S}(S_\tau^{t,z,\gamma}, Y_\tau^{t,z,\gamma}, \Theta_\tau^{t,z,\gamma}), Y_\tau^{t,z,\gamma} - \Theta_\tau^{t,z,\gamma})\ \right] < 1$$

where $z = (\mathcal{S}(s, y, -\theta), y + \theta, \theta, v)$.

Proof. The proof is analogous to [BLZ16, Proof of Prop.3.3]. We present it for completeness.

It is easy to see that for all $k \geq 2$ and $(t, s, y, \theta) \in [0, T] \times \mathbb{R}_+ \times \mathbb{R} \times (\mathcal{K} \cap [-k, k])$

$$\bar{w}_k(t, s, y, \theta) \geq w_{k+1}(t, \mathcal{S}(s, y, \theta), \mathcal{Y}(y, \theta)), \tag{4.18}$$

$$w_{k-1}(t, \mathcal{S}(s, y, \theta), \mathcal{Y}(y, \theta)) \geq \bar{w}_k(t, s, y, \theta). \tag{4.19}$$

Now suppose that $v > w(t, s, y)$. Then by definition of w there exists $\theta \in \mathcal{K}$ and some $\gamma \in \mathcal{G}(t, z)$ for $z = (\mathcal{S}(s, y, -\theta), y+\theta, \theta, v)$. As in [ST02, Proof of Thm.3.1, Step 1], we have for all stopping times $\tau \geq t$ (Part 1 of) the DPP for $\bar{w}$: $V_\tau^{\text{liq},t,z,\gamma} \geq \bar{w}(\tau, S_\tau^{t,z,\gamma}, Y_\tau^{t,z,\gamma}, \Theta_\tau^{t,z,\gamma})$;

we postpone the detailed proof for Section 4.8.1. Thus, Part 1 of the DPP for w follows from (4.18) by taking $k \to \infty$.

To prove Part 2, let $v < w_{2k+2}(t,s,y)$ and suppose that there exists $\gamma \in \Gamma_k$, $\theta \in \mathcal{K} \cap [-k,k]$ and a stopping time $\tau \geq t$ such that

$$V_\tau^{\mathrm{liq},t,z,\gamma} > w_k(\tau, \mathcal{S}(S_\tau^{t,z,\gamma}, Y_\tau^{t,z,\gamma}, \Theta_\tau^{t,z,\gamma}), Y_\tau^{t,z,\gamma} - \Theta_\tau^{t,z,\gamma})$$

for $z = (\mathcal{S}(s,y,-\theta), y+\theta, \theta, v)$. Then by (4.19) $V_\tau^{\mathrm{liq},t,z,\gamma} > \bar{w}_{k+1}(S_\tau^{t,z,\gamma}, Y_\tau^{t,z,\gamma}, \Theta_\tau^{t,z,\gamma})$ and thus, by [ST02, Proof of Thm.3.1, Step 2], we get that $v \geq \bar{w}_{2k+1}(t, \mathcal{S}(s,y,-\theta), y+\theta, \theta)$; the detailed arguments for the latter will be given in Section 4.8.1. In particular, by (4.18) we conclude that $v \geq w_{2k+2}(t,s,y)$, thus a contradiction. $\square$

Remark 4.3.2. The second part in the above theorem is stated in terms of w_k instead of w because of a measurable-selection argument employed in the proof, cf. Section 4.8.1.

To derive the pricing pde from the DPP in Theorem 4.3.1, we need the dynamics of the continuous processes

$$t \mapsto V_t^{\mathrm{liq}} - \varphi(t, \mathcal{S}(S_t, Y_t^\Theta, \Theta_t), \mathcal{Y}(Y_t^\Theta, \Theta_t)) \tag{4.20}$$

for sufficiently smooth functions $\varphi : [0,T] \times \mathbb{R}_+ \times \mathbb{R}$; they will later serve as test functions when characterizing w as a viscosity solution of a suitable pde.

Lemma 4.3.3. *For every $\gamma = (a,b,\nu) \in \Gamma$ and every $\varphi \in C^{1,2,1}([0,T] \times \mathbb{R}_+ \times \mathbb{R})$,*

$$\begin{aligned} &\mathrm{d}(V_t^{liq} - \varphi(t, \mathcal{S}_t, \mathcal{Y}_t)) = \\ &\qquad \mathcal{S}_t \left(\frac{F(\mathcal{Y}_t + \Theta_t) - F(\mathcal{Y}_t)}{f(\mathcal{Y}_t)} - \varphi_S \right) \{((\mu_t - \lambda(\mathcal{Y}_t) h(\mathcal{Y}_t + \Theta_t))\, \mathrm{d}t + \sigma\, \mathrm{d}W_t\} \\ &\qquad + \left\{ -\varphi_t - 1/2 \sigma^2 \mathcal{S}_t^2 \varphi_{SS} + h(\mathcal{Y}_t + \Theta_t) \varphi_Y + \mathfrak{F}(\mathcal{S}_t, \mathcal{Y}_t, \Theta_t) \right\} \mathrm{d}t, \end{aligned}$$

with

$$\mathfrak{F}(s,y,\theta) = sh(y+\theta) \left(\lambda(y) \frac{F(y+\theta) - F(y)}{f(y)} - \frac{f(y+\theta) - f(y)}{f(y)} \right),$$

where $\mathcal{S}_t = \mathcal{S}(S_t, Y_t^\Theta, \Theta_t)$, $\mathcal{Y}_t = \mathcal{Y}(Y_t^\Theta, \Theta_t)$ and the derivatives of φ are evaluated at $(t, \mathcal{S}_t, \mathcal{Y}_t)$.

Proof. Since $\mathcal{S}(S_t, Y_t^\Theta, \Theta_t) = \overline{S}_t f(Y_t^\Theta - \Theta_t)$, we get by the product rule (recall that $f' = \lambda f$)

$$\mathrm{d}\mathcal{S}_t = \mathcal{S}_t \left\{ (\mu_t - \lambda(Y_t^\Theta - \Theta_t) h(Y_t^\Theta))\, \mathrm{d}t + \sigma\, \mathrm{d}W_t \right\} \tag{4.21}$$

An application of Itô's formula gives

$$\begin{aligned}
\mathrm{d}\varphi(t,\mathcal{S}_t, Y_t^\Theta - \Theta_t) &= \varphi_t \,\mathrm{d}t + \varphi_S \,\mathrm{d}\mathcal{S}_t + \varphi_Y \,\mathrm{d}(Y_t^\Theta - \Theta_t) + 1/2\varphi_{SS} \,\mathrm{d}[\mathcal{S}]_t \\
&= \big\{\varphi_t - \lambda(Y_t^\Theta - \Theta_t)h(Y_t^\Theta)\mathcal{S}_t\varphi_S - h(Y_t^\Theta)\varphi_Y + 1/2\sigma^2\mathcal{S}_t^2\varphi_{SS}\big\} \,\mathrm{d}t \\
&\qquad + \mu_t\mathcal{S}_t\varphi_S \,\mathrm{d}t + \sigma\mathcal{S}_t\varphi_S \,\mathrm{d}W_t.
\end{aligned} \tag{4.22}$$

With reference to (4.9), we have

$$\begin{aligned}
\mathrm{d}V_t^{\mathrm{liq}} &= -h(Y_t^\Theta)\mathcal{S}_t \frac{f(Y_t^\Theta) - f(Y_t^\Theta - \Theta_t)}{f(Y_t^\Theta - \Theta_t)} \,\mathrm{d}t \\
&+ \mu_t\mathcal{S}_t \frac{F(Y_t^\Theta) - F(Y_t^\Theta - \Theta_t)}{f(Y_t^\Theta - \Theta_t)} \,\mathrm{d}t + \sigma\mathcal{S}_t \frac{F(Y_t^\Theta) - F(Y_t^\Theta - \Theta_t)}{f(Y_t^\Theta - \Theta_t)} \,\mathrm{d}W_t
\end{aligned} \tag{4.23}$$

Combining (4.22) and (4.23) and rearranging the terms finishes the proof. □

Remark 4.3.4. Consider the case when λ is constant, i.e. $f = \exp(\lambda\cdot)$, leading to the simplification $\mathfrak{F} \equiv 0$. In this case the dynamics of V^{liq} takes the following surprisingly simple form

$$\mathrm{d}V_t^{\mathrm{liq}} = F(\Theta_t) \,\mathrm{d}\mathcal{S}_t,$$

where $\mathcal{S}_t = \mathcal{S}(S_t, Y_t^\Theta, \Theta_t)$ has the dynamics (4.21). As a consequence, the minimal superhedging price (for the large investor) of an option with maturity T and *pure cash settlement* $H(S_T)$ is at least the small investor's price of H in absence of the large trader (when the price process is $\overline{S}$ instead). Indeed, for each superheding (for the large investor) strategy Θ with initial capital v there exists $\mathbb{P}^\Theta \approx \mathbb{P}$ (on $\mathcal{F}_T$) such that $\mathcal{S} = S_{0-}\mathcal{E}(\sigma\widetilde{W})$ under $\mathbb{P}^\Theta$, where $\widetilde{W}$ is a $\mathbb{P}^\Theta$-Brownian motion. Hence, $V^{\mathrm{liq}}(\Theta)$ is a $\mathbb{P}^\Theta$-martingale and thus $v \geq \mathbb{E}^{\mathbb{P}^\Theta}[H(S_T)] = \mathbb{E}^{\mathbb{P}^\Theta}[H(\mathcal{S}_T)]$ (recall that $\Theta_T = 0$ giving $S_T = \mathcal{S}_T$). On the other hand, the Feynman-Kac formula gives that $\mathbb{E}^{\mathbb{P}^\Theta}[H(\mathcal{S}_T)]$ is exactly the Black-Scholes price for a small investor in a market with risky asset process $\overline{S}$. Since Θ was arbitrary superhedging strategy with initial capital v, the infimum over all such strategies will again be bounded from below by the Black-Scholes price of the option.

Note that this is a notable difference to [BB04, Thm. 5.3], where the price for the large investor would be typically smaller. This is mainly due to their specification of superhedging strategies according to which a large trader would try to minimize at maturity the payoff of the option by exploiting his influence on the prices, i.e. he might change at maturity his risky asset position (and hence the price) in order to minimize the payoff, and immediately afterwards will unwind the difference at no additional cost (due to absence bid-ask spread). In contrast, we rule out such strategic behavior by imposing a constraint on the strategies to replicate exactly the physical delivery part.

At this point we should stress that similar argument does not extend to the general case of non-constant λ, and in fact we will see in Section 4.6 situations where it might be even cheaper for the large trader to superhedge the option, mainly due to the resilience effect, see also Example 4.6.1.

4.4 The pricing PDEs and main results

Let us first derive the T value of the value function w that will serve as a boundary condition for the pricing pde. Recall that $\mathcal{K}$ is the (constraint) set where the trading strategies are assumed to take values and set $\mathcal{K}_n := \mathcal{K} \cap [-n, n]$ for each $n \in \mathbb{N}$.

Lemma 4.4.1 (Boundary condition)**.** *Let for $n \in \mathbb{N}$*

$$H_n(s,y) := \inf \left\{ g_0\big(s\tfrac{f(y+\theta)}{f(y)}, y+\theta\big) + s\tfrac{F(y+\theta)-F(y)}{f(y)} \mid \theta \in \mathcal{K}_n,\ \theta = g_1\big(s\tfrac{f(y+\theta)}{f(y)}, y+\theta\big) \right\}.$$

Then, $w_n(T,\cdot) = H_n(\cdot)$ and $w(T,\cdot) = H(\cdot)$, where

$$H := \inf_{n\geq 0} H_n. \tag{\textbf{BC}}$$

Proof. At time T, the hedger of the option can do a block trade of size θ in order to meet the physical delivery part, moving the price from s to $s\frac{f(y+\theta)}{f(y)}$ and the impact from y to $y+\theta$. This block trade incurs costs of size $s\frac{F(y+\theta)-F(y)}{f(y)}$ and hence it superhedges the payoff (g_0, g_1) if the hedger can cover this costs and the cash delivery part, which after the block trade is $g_0\big(s\frac{f(y+\theta)}{f(y)}, y+\theta\big)$. □

Remark 4.4.2. Note that $H(s,y) = +\infty$ if the equation $\theta = g_1\big(s\frac{f(y+\theta)}{f(y)}, y+\theta\big)$ does not have a solution $\theta \in \mathcal{K}$.

Since at this point we don't know if the value function w is continuous, we need to work with discontinuous viscosity solutions and hence to consider the relaxed semi-limits

$$w_*(t,s,y) := \liminf_{(t',s',y',k)\to(t,s,y,\infty)} w_k(t',s',y'), \tag{4.24}$$

$$w^*(t,s,y) := \limsup_{(t',s',y',k)\to(t,s,y,\infty)} w_k(t',s',y'), \tag{4.25}$$

where the limits are taken over $t' < T$. Recall that w is a (discontinuous) viscosity solution of our pricing equations, see Sections 4.4.1 and 4.4.2, if w_* (resp. w^*) is a supersolution (resp. subsolution). To prove the viscosity property later, we need the following assumption.

Assumption 4.4.3.

Bounded value function: *w_* and w^* are bounded on $[0,T] \times \mathbb{R}_+ \times \mathbb{R}$;*

Regular payoff: *H is continuous, bounded, and $H_n \downarrow H$ uniformly on compacts.*

In particular, Assumption 4.4.3 implies that $w(T,\cdot)$ is finite. This means that the payoff is well-behaved in terms of the physical delivery part, i.e. the absurd situation from Remark 4.4.2 of having to deliver at maturity more risky assets than the market allows is ruled out.

4.4.1 The case of bounded impact function

In this section, we will work under the following assumption on the price impact function.

Assumption 4.4.4. *The resilience function h is Lipschitz and bounded, f is bounded away from 0 and ∞, i.e.* $\inf_{\mathbb{R}} f > 0$ *and* $\sup_{\mathbb{R}} f < +\infty$, *$\lambda$ is bounded and continuously differentiable with bounded derivative, and $\mathcal{K} = \mathbb{R}$, i.e. we do not impose delta constraints.*

Under Assumption 4.4.4, the antiderivative F from (4.7) (and its inverse F^{-1}) is a bijection on $\mathbb{R}$ and also Lipschitz continuous with Lipschitz constant $\sup_{\mathbb{R}} f < +\infty$ ($1/\inf_{\mathbb{R}} f$ respectively).

To derive the pricing pde in this case, let $(t, s, y) \in [0, T) \times \mathbb{R}_+ \times \mathbb{R}$ and apply formally Part 1 of the DPP in Theorem 4.3.1 to $v = w(t, s, y)$ (assuming that the infimum in the definition of w is attained) and $\tau = t+$, together with Lemma 4.3.3 for $\varphi = w$, assuming that w is smooth enough. Thus, we get the existence of θ^* such that

$$\begin{aligned} 0 \leq & s\big(\tfrac{F(y+\theta^*)-F(y)}{f(y)} - w_S(t,s,y)\big)\big\{(\mu_t - \lambda(y)h(y+\theta^*))\,\mathrm{d}t + \sigma\,\mathrm{d}W_t\big\} \\ & + \big\{ - w_t(t,s,y) - \tfrac{1}{2}\sigma^2 s^2 w_{SS}(t,s,y) + h(y+\theta^*)w_Y(t,s,y) + \mathfrak{F}(s,y,\theta^*)\big\}\,\mathrm{d}t. \end{aligned}$$

Still at a formal level, this cannot hold unless

$$\begin{aligned} F(y+\theta^*) = f(y)w_S(t,s,y) + F(y) \qquad \text{and} \\ -w_t(t,s,y) - 1/2\sigma^2 s^2 w_{SS}(t,s,y) + h(y+\theta^*)w_Y(t,s,y) + \mathfrak{F}(s,y,\theta^*) \geq 0. \end{aligned} \tag{4.26}$$

In particular, $\theta^* = \theta^*(t, y, s) = F^{-1}\big(f(y)w_S(t, s, y) + F(y)\big) - y$. The second part of DPP in Theorem 4.3.1 will actually give that the drift term must be 0, i.e. we should have equality in (4.26). This formally yields the following pde for w

$$-w_t - \frac{1}{2}\sigma^2 s^2 w_{SS} + \tilde{h}(t,s,y)\left(w_Y + s\lambda(y)w_S + s - s\tilde{f}(t,s,y)/f(y)\right) = 0, \quad (\mathbf{PDE})$$

where for $(t, s, y) \in [0, T) \times \mathbb{R}_+ \times \mathbb{R}$

$$\begin{aligned} \tilde{h}(t,s,y) &:= h \circ F^{-1}\big(f(y)w_S(t,s,y) + F(y)\big), \\ \tilde{f}(t,s,y) &:= f \circ F^{-1}\big(f(y)w_S(t,s,y) + F(y)\big). \end{aligned}$$

Indeed, our main result is

Theorem 4.4.5. *Under Assumption 4.4.3 and Assumption 4.4.4, the value function w is continuous and the unique bounded viscosity solution of* (**PDE**) *with the boundary condition $w(T, \cdot) = H(\cdot)$, where H is defined in* (**BC**)*.*

Proof. The viscosity property, i.e. that w_* (resp. w^*) is a viscosity supersolution (resp. subsolution), follows by the dynamic programming principle in Theorem 4.3.1 together with Lemma 4.3.3. The key arguments will be presented in Section 4.8 for the case when λ is constant, which would lead to the slightly more involved pricing pde ($\mathbf{PDE}^\delta$) (including gradient constraints) requiring additional justification. Moreover, in

Chapter 5 we will prove viscosity property of the value function for the superhedging problem in a multi-asset setup, cf. Propositions 5.3.10 and 5.3.11, arguments being very similar to these needed here.

The comparison result of Theorem 4.8.5 proves the rest of the claim, see also Remark 4.8.7. □

Let us conclude this section with some consequence of Theorem 4.4.5 for the minimal superhedging price and the existence of a minimal hedging strategy. A numerical example will be presented in Section 4.6.

Remark 4.4.6 (Dependence on displacement from unaffected price)**.** Like in the classical case of liquid markets (without market impact), the superhedging price does not depend on the drift in the unperturbed price process. This may be seen more directly by working under the equivalent martingale measure for $\overline{S}$ from the beginning. On the other hand, the superhedging price depends non-trivially on the level of impact y and the resilience function h, and can do so even for option payoffs of the form $(g_0(s), 0)$. i.e. not depending on the level of impact. So it turns out that for the pricing and hedging (cf. Remark 4.4.8) the perturbation of the market price from the 'unaffected' value is a relevant state variable.

Remark 4.4.7 (Only permanent impact)**.** Note that with only permanent impact, that is for $h \equiv 0$, (**PDE**) simplifies to the Black-Scholes pricing pde and hence the minimal superhedging price for the large trader is the Black-Scholes price for the option with payoff H.

Remark 4.4.8 (Replicating strategy)**.** Suppose that $w \in C_b^{1,3,1}([0,T] \times \mathbb{R}_+ \times \mathbb{R})$ solves the pricing pde (**PDE**) with the boundary condition $w(T,\cdot) = H(\cdot)$. Then for every $\varepsilon > 0$ a superhedging strategy with initial capital $w(0,s,y) + \varepsilon$ can be constructed as follows. Consider the self-financing strategy (β, Θ) with $\beta_{0-} = w(0,s,y) + \varepsilon$, $\Theta_0 = F^{-1}(f(y)w_S(0,s,y) + F(y)) - y$, meaning that a block trade of size $\Delta\Theta_0 = \Theta_0$ is performed at time 0, and

$$\Theta_t = F^{-1}\left(f(\mathcal{Y}_t^\Theta)w_S(t, \mathcal{S}(S_t, Y_t^\Theta, \Theta_t), \mathcal{Y}_t^\Theta) + F(\mathcal{Y}_t^\Theta)\right) - \mathcal{Y}_t^\Theta \quad \text{for } t \in [0,T), \tag{4.27}$$

$$\Theta_T = 0, \quad \text{i.e.} \quad \Delta\Theta_T = \Theta_{T-}, \tag{4.28}$$

where $\mathcal{Y}^\Theta = Y^\Theta - \Theta$. Then by Lemma 4.3.3, together with (4.27) and (**PDE**) we conclude that

$$\begin{aligned}
\varepsilon &= V_0^{\mathrm{liq}}(\Theta) - w(0,s,y) = V_T^{\mathrm{liq}}(\Theta) - w(T, \mathcal{S}(S_T, Y_T^\Theta, \Theta_T), \mathcal{Y}_T^\Theta) \\
&= V_T^{\mathrm{liq}}(\Theta) - H(\mathcal{S}(S_T, Y_T^\Theta, \Theta_T), \mathcal{Y}_T^\Theta) \\
&= V_T^{\mathrm{liq}}(\Theta) - H(S_T, Y_T^\Theta), \qquad \Theta_T = 0,
\end{aligned}$$

where the last line follows from (4.28). By definition of H, $H + \varepsilon$ we will be enough to superreplicate the European claim with payoff (g_0, g_1) with a possible additional block trade of size Δ^ε at time T (note that such a block trade will not affect V_T^{liq}). Hence, the strategy $\Theta + \mathbb{1}_{\{T\}}\Delta^\varepsilon$ will be superreplicating for the European claim with payoff H. Note that one could take $\varepsilon = 0$ if the constructed strategy is bounded and the infimum in the definition of H_n is attained (cf. Lemma 4.4.1), i.e. we have a replicating strategy in this case.

An application of Itô's formula gives that a strategy Θ satisfying the fixed-point problem (4.27) can be obtained by considering the following system of SDEs:

$$\begin{aligned} \mathrm{d}\mathcal{S}_t &= \mathcal{S}_t[(\mu_t - \lambda(\mathcal{Y}_t^\Theta)h(\mathcal{Y}_t^\Theta + \Theta_t))\,\mathrm{d}t + \sigma\,\mathrm{d}W_t], \\ \mathrm{d}\Theta_t &= a(t, \mathcal{S}_t, \mathcal{Y}_t^\Theta, \Theta_t)\,\mathrm{d}t + b(t, \mathcal{S}_t, \mathcal{Y}_t^\Theta)\,\mathrm{d}W_t, \\ \mathrm{d}\mathcal{Y}_t^\Theta &= -h(\mathcal{Y}_t^\Theta + \Theta_t)\,\mathrm{d}t, \end{aligned} \tag{4.29}$$

with initial conditions $\mathcal{S}_0 = s$, $\mathcal{Y}_0^\Theta = y$ and $\Theta_0 = F^{-1}(f(y)w_S(0, s, y) + F(y)) - y$, where (with $f = f(y), \lambda = \lambda(y)$ etc. for ease of notation)

$$\begin{aligned} a(t, s, y, \theta) &:= h(y + \theta)\left(1 - \frac{\lambda f w_S - f - w_{SY} - \lambda s w_{SS}}{f(F^{-1}(f w_S + F))}\right) + \\ &\qquad + \frac{w_{tS} + s\mu_t w_{SS} + 1/2\sigma^2 s^2 w_{SSS}}{f(F^{-1}(f w_S + F))}, \\ b(t, s, y) &:= \frac{\sigma s w_{SS}}{f(F^{-1}(f w_S + F))}. \end{aligned}$$

Hence, an optimal superhedging strategy will also account for the transient nature of price impact.

We close this section with a remark on Assumption 4.4.4 that implies the bijectivity of F on $\mathbb{R}$. In particular, this ensures that an optimal control θ^* can be defined. Similar conditions are also crucial for the results in [BB04] and [BLZ16], namely the surjectivity assumption A5 in [BB04] and the invertability assumption H2 in [BLZ16]. We will see in the next section how departing from this assumption leads naturally to singularity in the pricing pde with respect to the gradient. Indeed, the lack of invertability of F imposes a condition on w_S so that θ^* could be defined. Thus, for the analysis there we will introduce constraints on the "delta", i.e. the holdings in the risky asset, which in pde terms translates to constraints on the spacial gradient w_S.

4.4.2 The case of exponential impact function

In this section, we consider the case of price impact $f(x) = \exp(\lambda x)$ being exponential, meaning that the relative marginal price impact function $\lambda = f'/f > 0$ is constant. A peculiarity of this case is that at any time instant t, knowing the (marginal) price S_t for the stock is sufficient to know the impact from an instant block trade, since after a block trade of size Δ the price would be $\overline{S}_t f(Y_t + \Delta) = S_t \exp(\lambda\Delta)$. Hence, the relative displacement $f(Y^\Theta)$ of S from the fundamental price $\overline{S}$ is immaterial to determine the price impact from a block trade, in difference to the situation of Section 4.4.1. Motivated by Remark 4.2.5, we consider trading with short-selling constraints, i.e. trading strategies are required to take values in $\mathcal{K} = [-K, \infty)$ for some $K > 0$.

To derive (heuristically, at first) the pricing pde, let us apply formally Theorem 4.3.1 for $v = w(t,s,y)$ at $t, s, y, \tau = t+$, provided that w is smooth enough, to get the existence of $\theta^* \in \mathcal{K}$ such that, using Lemma 4.3.3, we have

$$\mathcal{L}^{\theta^*} w(t,s,y)\,\mathrm{d}t - s(w_S(t,s,y) - e^{\lambda\theta^*}/\lambda + 1/\lambda)(\sigma\,\mathrm{d}W_t + \eta_t\,\mathrm{d}t) \geq 0,$$

where $\eta_t = \mu_t - \lambda h(y+\theta^*)$ and

$$\mathcal{L}^{\theta^*} w(t,s,y) := -w_t(t,s,y) + h(y+\theta^*)w_Y(t,s,y) - \frac{1}{2}\sigma^2 s^2 w_{SS}(t,s,y).$$

As in Section 4.4.1, the diffusion part should vanish, giving the optimal control

$$\theta^* = \frac{1}{\lambda}\log\big(\lambda w_S(t,s,y) + 1\big),$$

and from the drift part we identify the pricing pde $\mathcal{L}^{\theta^*} w(t,s,y) = 0$. The constraint $\theta^* \in \mathcal{K}$ is now equivalent to $\mathcal{H}_{\mathcal{K}} w(t,s,y) \geq 0$, where for a smooth function φ

$$\mathcal{H}_{\mathcal{K}}\varphi(t,s,y) := \lambda\varphi_S(t,s,y) + 1 - e^{-\lambda K}$$

Thus we obtain, just formally, that w should be a solution to the variational inequality

$$\mathcal{F}_{\mathcal{K}}[w] := \min\{\mathcal{L}^{\theta[w]}w\ ,\ \mathcal{H}_{\mathcal{K}}w\} = 0 \quad \text{on } [0,T) \times \mathbb{R}_+ \times \mathbb{R}, \tag{$\mathbf{PDE}^\delta$}$$

where

$$\theta[w](t,s,y) := 1/\lambda \cdot \log\big(\lambda w_S(t,s,y) + 1\big). \tag{4.30}$$

It turns out that the gradient constraints $\mathcal{H}_{\mathcal{K}} w \geq 0$ on the value function, that hold on $[0,T)$, propagate to the boundary, meaning that the correct boundary condition for ($\mathbf{PDE}^\delta$) is

$$\min\{w(T,\cdot) - H, \mathcal{H}_{\mathcal{K}}w\} = 0. \tag{$\mathbf{BC}^\delta$}$$

Next we state our main result for exponential price impact function.

Theorem 4.4.9. *Suppose that the resilience function h is Lipschitz continuous and Assumption 4.4.3 is in force. Then the minimal superhedging price w of an European option with maturity T and payoff profile (g_0, g_1) is the unique bounded viscosity solution of the variational inequality ($\mathbf{PDE}^\delta$) with boundary condition ($\mathbf{BC}^\delta$). In particular, $w_* = w^* = w$ on $[0,T] \times \mathbb{R}_+ \times \mathbb{R}$.*

Proof. The proofs are postponed for Section 4.8. The viscosity super-/sub-solution property are proved in Theorem 4.8.2 and Theorem 4.8.3 respectively, while uniqueness follows from the comparison result Theorem 4.8.6, see also Remark 4.8.7. □

Corollary 4.4.10. *In the setup from Theorem 4.4.9, suppose moreover that the payoff (g_0, g_1) does not depend on the level of impact y. Then the minimal superhedging price*

is a function in (t,s) *only and the pricing pde* (4.30) *simplifies to the Black-Scholes pde with gradient constraints.*
In this case, if the face-lifted payoff $F_{\mathcal{K}}[H]$ *is continuously differentiable with bounded derivative, where*

$$F_{\mathcal{K}}[H](s) := \sup_{x\leq 0}\left\{H(s+x) + \frac{1-e^{-\lambda K}}{\lambda}x\right\}, \quad s \in \mathbb{R}_+,$$

with the convention that $H = H(0)$ *on* $(-\infty, 0]$, *then the minimal superhedging price coincides with the Black-Scholes price for the face-lifted payoff* $F_{\mathcal{K}}[H]$.

Proof. If (g_0, g_1) is a function of the price process s only, then it is easy to see that H is such as well and that the dimension of the state process can be reduces by ignoring the impact process Y. In this case, the stochastic target problem in Section 4.3 could be formulated for the new state process and thus the value function would be a function on (t,s) only. The same analysis could be carried over to derive the pricing pde and to prove viscosity solution property of the value function. The pricing pde in this case would be the Black-Scholes pde with gradient constraints since the term $h(Y)\varphi_Y$ in Lemma 4.3.3 would not be present. Hence, the minimal superhedging price in our large investor model would coincide with the minimal superhedging price under delta constraints in the small investor model for the payoff H (because it solves the same pde). In this one-dimensional setup, this price coincides with the Black-Scholes price for the face-lifted payoff $F_{\mathcal{K}}[H]$, cf. [CEK15, Proposition 3.1]. □

Example 4.4.11. Consider the contingent claim with payoff $H(s) = s$. In the friction-less Black-Scholes world, the present value of this claim is the price of the underlying, simply because $w^{\mathrm{BS}}(t,s) = s$ solves the Black-Scholes pde with the terminal condition H, and the replicating strategy in this case consists of holding $w^{\mathrm{BS}}_S(t,s) = 1$ asset, i.e. it is a buy-and-hold strategy. To see that a buy-and-hold strategy is also optimal for the large trader when $f(y) = \exp(\lambda y)$, even in the case with transient price impact, note that for initial capital s in the riskless asset and impact level y, the large trader could buy at the beginning with an immediate block trade exactly $\theta(s,y) = 1/\lambda\log(1+\lambda)$ shares. The key property in the case of exponential f is that p does not depend on s and y. Hence, after buying $\theta(s,y)$ shares and holding them until maturity T, where the new price and impact would be S_{T-} and Y_{T-} respectively, the large trader performs a block trade to unwind his risky asset position and receives exactly

$$\overline{S}_T(F(Y_{T-}) - F(Y_{T-}-\theta)) = \overline{S}_T\exp(\lambda(Y_{T-}-\theta))\frac{\exp(\lambda\theta)-1}{\lambda} = \overline{S}_T f(Y_{T-}-\theta) = S_T$$

in cash, where S_T will be the price after the final liquidation block trade. Hence, with this buy-and-hold strategy of $1/\lambda\log(1+\lambda)$ shares, that requires exactly capital s at the beginning, the large trader will be able to replicate the claim with payoff H. Moreover, the arguments in Remark 4.3.4 show that the minimal initial capital for the large trader in this case cannot not be less that in the friction-less case, hence we just constructed an optimal hedging strategy for the large trader.

Quite surprisingly at first sight, the large trader does not exploit the level of impact and possible drift that the resilience of price impact creates. The reason in this simple case of exponential f is that the level of impact is irrelevant in determining the additional cost of performing a block trade. This would no longer be the case in the setup of Section 4.4.1 where the level of impact also determines the cost of a block trade. In addition, our superreplication approach forces the trader to take early actions and not exploit directly the possible drift due to resilience of impact. Indeed, imposing the almost sure constraint at maturity forces the large trader to trade earlier and undertake a non-zero position in the risky asset in order to hedge away the endogenous risks. We expect that even in this simple example of exponential f, another hedging approach that relaxes the almost sure constraints at maturity, e.g. utility indifference pricing or mean-variance hedging, could underline even more the role of transient impact on pricing and hedging. This will be left for future studies.

4.5 Combined transient and permanent price impact

The price impact considered so far is purely transient, meaning that wears off over time. Here we show how our analysis can be extended to intertemporal impact which has a transient and also a permanent component. To this end, we can modify our model as follows.

For $\eta \geq 0$ let the price of the risky asset for infinitesimal trade be given by the following modification of (4.3):

$$S_t = f(\eta\Theta_t + Y_t^{\Theta})\overline{S}_t, \tag{4.31}$$

where Y^{Θ} is given by (4.2). If the large trader is inactive, the price process will recover towards $f(\eta\Theta)\overline{S}$ due to the mean-reversion property of Y^{Θ}. Adapting the analysis from Chapter 2 to the current setup, see also the discussion in Section 2.4.4, we obtain the asymptotically realizable proceeds from a general semimartingale strategy Θ:

$$\tilde{L}(\Theta) = \frac{1}{1+\eta}\left(\int_0^{\cdot} F(\eta\Theta_t + Y_t^{\Theta})\,\mathrm{d}\overline{S}_t - \int_0^{\cdot} \overline{S}_t f(\eta\Theta_t + Y_t^{\Theta}) h(Y_t^{\Theta})\,\mathrm{d}t - \overline{S}F(\eta\Theta + Y^{\Theta})\big|_{0-}^{\cdot}\right).$$

In particular, a block trade $\Delta\Theta_t$ yields $-\overline{S}_t \frac{1}{1+\eta}\int_0^{(1+\eta)\Delta\Theta_t} f(\eta\Theta_{t-} + Y_{t-}^{\Theta} + x)\,\mathrm{d}x$ in proceeds. Thus, following the discussion in Section 4.1 the volume effect process (in the spirit of [PSS11]) in this case is $\eta\Theta + Y^{\Theta}$ and so the volume imbalances from trading also have a permanent component.

The instantaneous liquidation value process $\tilde{V}^{\mathrm{liq}}$ in this modified model now satisfies

$$(1+\eta)\,\mathrm{d}\tilde{V}_t^{\mathrm{liq}} = (F(\eta\Theta_t + Y_t^{\Theta}) - F(Y_t^{\Theta} - \Theta_t))\,\mathrm{d}\overline{S}_t - h(Y_t^{\Theta})(f(\eta\Theta_t + Y_t^{\Theta}) - f(Y_t^{\Theta} - \Theta_t))\,\mathrm{d}t.$$

Note that this way of incorporating permanent impact will not affect the effective price and impact processes $\mathcal{S}(S, Y^{\Theta}, \Theta)$ and $\mathcal{Y}(Y^{\Theta}, \Theta)$ since the permanent component vanishes for zero shares in the risky asset. This in turn gives that the same analysis so far can be carried over also here, with the following slight adjustments being needed:

- the boundary condition in Lemma 4.4.1 needs to be modified by adding the prefactor $1+\eta$ of θ, when an argument of a function,
- in Lemma 4.3.3, $F(Y^\Theta)$ should be substituted by $F(\eta\Theta + Y^\Theta)$, all the fractions should be divided by $1+\eta$ and $\mathfrak{F}$ should be replaced by $\mathfrak{F}^\eta$ with

$$\mathfrak{F}^\eta(s,y,\theta) := sh(y+\theta)\left(\lambda(y)\tfrac{F(y+(1+\eta)\theta)-F(y)}{(1+\eta)f(y)} - \tfrac{f(y+(1+\eta)\theta)-f(y)}{(1+\eta)f(y)}\right).$$

Let us first discuss the setup of Section 4.4.1 which essentially required F to be invertible on $\mathbb{R}$. In this case, the pricing pde will have the same structure as (**PDE**) with the following modifications: replace $\tilde{h}$ and $\tilde{f}$ by $\tilde{h}^\eta$ and $\tilde{f}^\eta$ respectively with

$$\tilde{h}^\eta(t,s,y) = h\left(\tfrac{1}{1+\eta}F^{-1}((1+\eta)f(y)\varphi_S(t,s,y)+F(y)) + \tfrac{\eta}{1+\eta}y\right),$$

$$\tilde{f}^\eta(t,s,y) = f\circ F^{-1}((1+\eta)f(y)\varphi_S(t,s,y)+F(y)).$$

An optimal hedging strategy Θ^*, if it exists, would satisfy (cf. Remark 4.4.8 for $\eta = 0$)

$$(1+\eta)\Theta^*_t = F^{-1}((1+\eta)f(\mathcal{Y}^*_t)\varphi_S(t,\mathcal{S}^*_t,\mathcal{Y}^*_t)+F(\mathcal{Y}^*_t)) - \mathcal{Y}^*_t,$$

where $\mathcal{S}^* = \mathcal{S}(S, Y^{\Theta^*}, \Theta^*)$ and $\mathcal{Y}^* = \mathcal{Y}(Y^{\Theta^*}, \Theta^*)$. Hence, the large trader's optimal strategy also reflects the permanent component in addition to the displacement from the "unaffected" price process tracked by Y^Θ.

In the setup of Section 4.4.2, we again need to consider portfolio constraints $\theta \in \mathcal{K}$ for $\mathcal{K} = [-K, +\infty)$ in order to derive the pricing pde. Moreover, $\mathfrak{F}^\eta = 0$ and thus the pricing pde takes the following form: $\forall(t,s,y) \in [0,T)\times\mathbb{R}_+\times\mathbb{R}$

$$\min\{-w_t - \tfrac{1}{2}\sigma^2 s^2 w_{SS} + h(y+\theta^*)w_Y,\ \lambda(1+\eta)\varphi_S + 1 - e^{-\lambda(1+\eta)K}\} = 0,$$

where $\theta^* = \frac{1}{\lambda(1+\eta)}\log\big(\lambda(1+\eta)w_S + 1\big)$, with boundary condition

$$\min\{w(T,\cdot) - H,\ \lambda(1+\eta)\varphi_S + 1 - e^{-\lambda(1+\eta)K}\} = 0,$$

where H is the modified boundary condition from Lemma 4.4.1 as explained above. In particular, the pricing pde with permanent as well as transient impact coincides with the pricing pde with pure transient impact but with modified λ, in this case $\lambda(1+\eta)$.

4.6 Numerical example

In this section, we discuss numerical results on the minimal superhedging price w characterized by (**PDE**), cf. Theorem 4.4.5. For our numerical simulations we consider impact function

$$f(x) = 1 + \arctan(x)/10, \quad x \in \mathbb{R}, \tag{4.32}$$

that satisfies Assumption 4.4.4. In this case the changes of $\lambda(x) = 1/(10(1+x^2)f(x))$ are most significant in $(-4, 4)$ where the change in impact is significant, see Figure 4.1a. Apart from satisfying our assumptions and having explicit antiderivative $F(x) = x + (x \arctan(x) - 1/2 \log(1 + x^2))/10$, being useful in the numerical implementation, it turns out that similar shape of the impact function was observed when the related Propagator model was calibrated to real data, see [BL12, Appendix] for details.

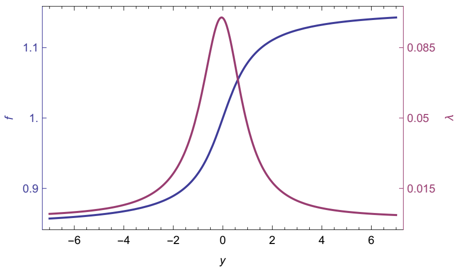
For $h(y) = \beta y$ with $\beta = 1$, we compare the large trader's price of a European call option with physical delivery at maturity $T = 0.5$ and strike $K = 50$, and its Black-Scholes price, i.e. the Black-Scholes price of a European call option for the same model parameters; let us recall that the case $f = 1$ in our market impact model gives the Black-Scholes model. The volatility σ is set to 0.3. The payoff for the large trader is $H(s, y) = \left(s \frac{F(y+1)-F(y)}{f(y)} - K \right) \mathbb{1}_{\{s \geq K\}}$ that we "smooth out" by approximating the indicator function by linearly interpolating 0 and 1 between $K - 0.5$ and K.

To approximate both prices, we solve the corresponding pdes using (semi-implicit) finite difference scheme in the bounded region $(y, s) \in [-20, 20] \times [0, 200]$. For our simulation we set the following boundary condition for $t < T$: $\frac{\partial w}{\partial s} = (F(y+1) - F(y)) / f(y)$ on $[-20, 20] \times \{200\}$, $\frac{\partial w}{\partial y} = 0$ on $\{-20, 20\} \times [0, 200] \cup [-20, 20] \times \{0\}$. Indeed, for initial impact y close to -20 or 20 the impact function is approximately constant and until maturity T resilience will not be able bring back the level of impact to the region where the changes in f are significant, see Figure 4.1a, thus we should expect that the price would not depend that much on the level of impact. On the other hand, for larger values of s one expects the price to depend linearly in s (like the payoff profile). The difference between the Black-Scholes price and the large trader's price (as a function of the risky asset price s and the level of impact y) is shown in Figure 4.1b. Let us point out that the Black-Scholes price does not depend on level of impact y.

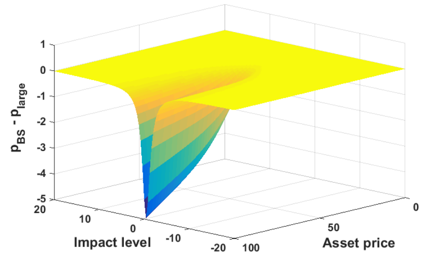
Although our numerical results suggest that the value of the option with physical delivery in our large trader model dominates its Black-Scholes price, this does not seem to be the case for the European call with pure cash delivery. In this case numerical simulations show that the price for the large investor can be smaller, typically when the initial impact is away from zero, i.e. in regions where the level of impact affects significantly the price when trading. The intuition is that for pure cash delivery, the net number of traded assets for a (super-)hedging strategy is zero (recall that $\Theta_{0-} = \Theta_T = 0$ for a superhedging Θ), while the presence of resilience incurs additional drift that could be favorable for the large trader, typically pushing the prices down if the hedging strategy consists of holding positive number of risky assets (and initial displacement is not small); see also Figure 4.1d.

On the other hand, superhedging becomes more expensive for the large trader when at maturity he has to deliver physically the asset, since at maturity he should have bought one asset (when the option is in-the-money) triggering price changes in unfavorable for his direction due to impact. In addition, we see that the presence of resilience renders the displacement from the fundamental price (the level of impact) an important new state variable.

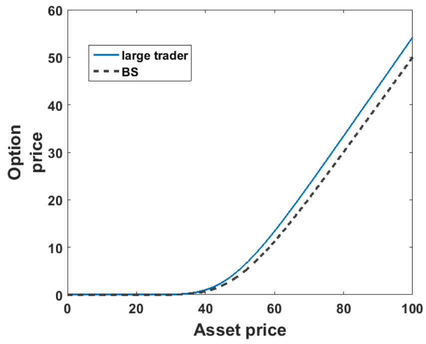
Example 4.6.1. In this example, we will show that the price of a European option



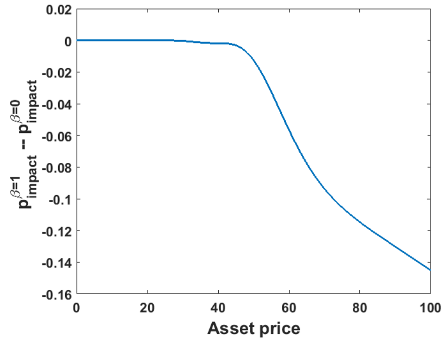
(a) Impact function f (in blue) and its logarithmic derivative λ (in purple)



(b) $p_{\mathrm{BS}} - p_{\mathrm{physical}}$ for resilience rate $\beta = 1$



(c) The Black-Scholes price and the large trader's price for call option with physical delivery, resilience rate $\beta = 1$ and initial impact level $y = 0$



(d) Difference between large trader's price for call option with physical delivery with and without resilience ($\beta = 1$ and $\beta = 0$ resp.), for initial impact level $y = 0$

Figure 4.1: Numerical simulations with impact function f from (4.32), $\sigma = 0.3$, $T = 0.5$, strike $K = 50$, resilience function $h(y) = \beta y$

in the Black-Scholes model (for the small investor) might indeed be greater than the minimal superhedging price for the large trader of this option with pure cash delivery. More specifically, for maturity $T > 0$ consider the solution v^{BS} of the Black-Scholes pde with bounded and smooth terminal condition H that has bounded derivatives, where we moreover assume that $\partial_S H \geq 0$, for instance a smooth approximation of a bull call spread option. Note that in particular $\partial_S v^{\mathrm{BS}} \geq 0$ and the derivatives of v^{BS} are bounded. We compare now $v^{\mathrm{BS}}(0, \cdot)$ with $v(0, \cdot, y)$ for large values of y, where $v = w$ with w from Theorem 4.4.5 with terminal condition H. Note that when $y = Y_{0-} > 0$ the affected price process includes additional drift in favorable for the large trader direction.

Let Θ with $\Theta_{0-} = 0$ be such that $\Theta_T = 0$ (corresponding to cash delivery at maturity) and for $t \in [0, T-]$

$$\Theta_t = F^{-1}(\partial_S v^{\mathrm{BS}}(t, \mathcal{S}_t) f(\mathcal{Y}_t^{\Theta}) + F(\mathcal{Y}_t^{\Theta})) - \mathcal{Y}_t^{\Theta}, \tag{4.33}$$

where $\mathcal{Y}^{\Theta} = Y^{\Theta} - \Theta$ and $\mathcal{S} = f(\mathcal{Y}^{\Theta})\overline{S}$. Since v^{BS} is smooth, the arguments in Remark 4.4.8

ensure existence of such Θ, while positivitity of $\partial_S v^{\mathrm{BS}}$ implies $\Theta \geq 0$ on $[0,T]$. Now for the self-financing portfolio (β,Θ) with initial cash holdings $\beta_{0-} = v^{\mathrm{BS}}(0,S_{0-})$ we have by (4.21), (4.23) and (4.33) (recall that $S_{0-} = \mathcal{S}_0$)

$$\begin{aligned} V_T^{\mathrm{liq}} &= v^{\mathrm{BS}}(0,\mathcal{S}_0) + \int_0^T \partial_S v^{\mathrm{BS}}(t,\mathcal{S}_t)\,\mathrm{d}\mathcal{S}_t \\ &\quad - \int_0^T \mathcal{S}_t h(Y_t^\Theta)\left(\frac{f(Y_t^\Theta) - f(Y_t^\Theta - \Theta_t)}{F(Y_t^\Theta) - F(Y_t^\Theta - \Theta_t)} - \lambda(Y_t^\Theta - \Theta_t)\right)\mathrm{d}t \\ &= H(\mathcal{S}_T) - \int_0^T \mathcal{S}_t h(Y_t^\Theta)\left(\frac{f(Y_t^\Theta) - f(Y_t^\Theta - \Theta_t)}{F(Y_t^\Theta) - F(Y_t^\Theta - \Theta_t)} - \lambda(Y_t^\Theta - \Theta_t)\right)\mathrm{d}t. \end{aligned} \tag{4.34}$$

In particular, if the integrand in (4.34) is negative on $[0,T]$, then (β,Θ) would be a superhedging strategy for the large trader with initial capital $\beta_{0-} = v^{\mathrm{BS}}(0,S_{0-})$ and hence

$$v(0,S_{0-},Y_{0-}) \leq v^{\mathrm{BS}}(0,S_{0-}). \tag{4.35}$$

One could show that the integrand will be negative for instance when $Y^\Theta \geq 0$ on $[0,T]$ and λ is strictly decreasing (at least on a compact set containing the range of Y^Θ and $Y^\Theta - \Theta$). Such a situation could arise if for example Y_{0-} is large enough. However, this should be intuitively favorable for the large trader due to the additional negative drift in the price that would suppress the underlying and hence the payout at maturity, i.e. in this case it could be expected that superhedging for the large trader might be cheaper. Let us point out that equality in (4.35) cannot hold for all values of S_{0-}, Y_{0-} since the two functions are solutions of different pdes; for non-constant λ, (**PDE**) does not simplify to the Black-Scholes pde like in Section 4.4.2.

4.7 The case of covered options

A key conclusion from [BLZ16, BLZ17] is that the way the hedger forms the hedging strategy and delivers the payoff is crucial for the pricing equation. In our setup so far the initial and terminal actions of the hedger have impact on the price and the minimal superhedging price is characterized by a semi-linear pde. We consider now the case of *covered options*, that is when the buyer of the option could be asked to provide the required initial hedging position and to accept a mix of cash and stocks (at their current market price) as a final payment, thus allowing the hedger of the option to escape initial and terminal impact of forming and unwinding the hedging position respective. The pricing equation turns out to be fully non-linear and degenerate in the second-order term. Since this is not our main contribution of this chapter, we restrict ourselves to a sketch of the derivation of the pricing pde and how one could adapt directly the analysis from [BLZ16].

Let us consider continuous hedging strategies that are Itô processes

$$\mathrm{d}\Theta_t = a(t)\,\mathrm{d}t + b(t)\,\mathrm{d}W_t, \qquad \Theta_0 = \theta_0 \in \mathbb{R}, \tag{4.36}$$

where a and b are continuous processes with some integrability conditions. For such controls Θ, the market impact process and the perturbed price process take the form:

$$\begin{aligned}
\mathrm{d}Y_t &= (-h(Y_t) + a(t))\,\mathrm{d}t + b(t)\,\mathrm{d}W_t, \quad Y_0 = y \\
\mathrm{d}S_t &= \mathrm{d}(f(Y_t)\bar{S}_t) \\
&= S_t\big[\underbrace{(\mu - \lambda(Y_t)h(Y_t) + \lambda(Y_t)a(t) + 0.5(\lambda(Y_t)^2 + \lambda'(Y_t))b^2(t) + \lambda(Y_t)\sigma b(t))}_{=:\xi(t)}\,\mathrm{d}t \\
&\qquad + (\sigma + \lambda(Y_t)b(t))\,\mathrm{d}W_t\big],
\end{aligned} \tag{4.37}$$

with $S_0 = f(y)\bar{s}$ (initial impact of acquiring θ_0 shares in the beginning is omitted). Note that θ_0 needs to be determined for a replicating strategy.

Performing integration by parts in (4.6), we can rewrite the gains from trading for continuous strategy Θ as

$$L_T(\Theta) = -\int_0^T S_t\,\mathrm{d}\Theta_t - \frac{1}{2}[S,\Theta]_T - \frac{1}{2}\int_0^T \sigma S_t\,\mathrm{d}[\Theta, W]_t \tag{4.38}$$

For a self-financing strategy (β, Θ) the book wealth at time T is (recall (4.10))

$$V_T^{\text{book}}(\Theta) = \beta_0 + L_T(\Theta) + \Theta_T S_T.$$

Consider a contingent claim of the form $H = g(S_T)$ written on the risky asset price. For a superhedging strategy Θ with initial capital p, the hedger needs to set up the initial position in the risky asset Θ_0 that incurs the cost ΘS_0. Hence, at maturity we have

$$p + L_T(\Theta) + \Theta_T S_T - \Theta_0 S_0 \geq g(S_T). \tag{4.39}$$

Using (4.38), the change in the book wealth satisfies

$$\begin{aligned}
\mathrm{d}L_t + \mathrm{d}(S_t\Theta_t) &= \Theta_t\,\mathrm{d}S_t + \frac{1}{2}\,\mathrm{d}[S,\Theta]_t - \frac{1}{2}\sigma S_t\,\mathrm{d}[\Theta, W]_t \\
&= \Theta_t\,\mathrm{d}S_t + \frac{1}{2}S_t\lambda(Y_t)b^2(t)\,\mathrm{d}t
\end{aligned} \tag{4.40}$$

Therefore, a replicating strategy Θ with initial capital p should satisfy

$$p + \int_0^T \Theta_t\,\mathrm{d}S_t + \int_0^T \frac{1}{2}S_t\lambda(Y_t)b^2(t)\,\mathrm{d}t = g(S_T).$$

To construct a replicating strategy we look for a pair of processes (a, b) (or equivalently a strategy Θ) such that the process $G_t := G_0 + \int_0^t \Theta_u\,\mathrm{d}S_u + \int_0^t \frac{1}{2}S_u\lambda(Y_u)b^2(u)\,\mathrm{d}u$ satisfies

$G_T = g(S_T)$. To find such a process, we try the following *Ansatz:* $G_t = v(t, S_t)$ for a smooth enough function $v : (0, \infty) \times \mathbb{R} \to \mathbb{R}$. Applying Itô's formula we get

$$\begin{aligned} \mathrm{d}v(t, S_t) &= v_t(t, S_t)\,\mathrm{d}t + v_s(t, S_t)\,\mathrm{d}S_t + \tfrac{1}{2}v_{ss}(t, S_t)\,\mathrm{d}[S]_t \\ &= [v_t + \tfrac{1}{2}S_t^2(\sigma + \lambda(Y_t)b(t))^2 v_{ss} + S_t\xi(t)v_s]\,\mathrm{d}t \\ &\qquad + S_t(\sigma + \lambda(Y_t)b(t))v_s(t, S_t)\,\mathrm{d}W_t. \end{aligned}$$

Comparing the drift and the diffusion terms we need

$$v_t + 1/2S_t^2(\sigma + \lambda(Y_t)b(t))^2 v_{ss} + v_s S_t\xi(t) = \frac{1}{2}S_t\lambda(Y_t)b^2(t) + \Theta_t S_t\xi(t), \tag{4.41}$$

$$(\sigma + \lambda(Y_t)b(t))S_t v_s(t, S_t) = \Theta_t S_t(\sigma + \lambda(Y_t)b(t)). \tag{4.42}$$

Now (4.42) is satisfied for $\Theta_t = v_s(t, S_t)$ and for this choice of Θ, (4.41) reduces to

$$v_t(t, S_t) + \tfrac{1}{2}S_t^2(\sigma + \lambda(Y_t)b(t))^2 v_{ss}(t, S_t) - \tfrac{1}{2}S_t\lambda(Y_t)b^2(t) = 0. \tag{4.43}$$

To get the form of b, we have by Itô's formula

$$\begin{aligned} a(t)\,\mathrm{d}t + b(t)\,\mathrm{d}W_t = \ &\mathrm{d}\Theta_t = \ \mathrm{d}v_s(t, S_t) = \\ &= v_{st}\,\mathrm{d}t + v_{ss}\,\mathrm{d}S_t + 1/2v_{sss}\,\mathrm{d}[S]_t \end{aligned}$$

and comparing the diffusion coefficients we get that $b(t) = v_{ss}S_t(\sigma + \lambda(Y_t)b(t))$, i.e.

$$b(t) = \frac{\sigma S_t v_{ss}(t, S_t)}{1 - \lambda(Y_t)S_t v_{ss}(t, S_t)}. \tag{4.44}$$

Similarly, we get

$$a(t) = v_{st} + v_{ss}S_t\xi(t) + \tfrac{1}{2}v_{sss}S_t^2(\sigma + \lambda(Y_t)b(t))^2.$$

Using the definition of ξ in (4.37), we get (with λ and λ' evaluated at Y_t)

$$a(t) = \frac{v_{st} + v_{ss}S_t[\mu - \lambda h(Y_t) + 0.5(\lambda^2 + \lambda')b^2(t) + \lambda\sigma b(t)] + 1/2v_{sss}S_t^2(\sigma + \lambda b(t))^2}{1 - \lambda S_t v_{ss}}.$$

Note that $\sigma + \lambda(Y_t)b(t) = \sigma/(1 - \lambda(Y_t)S_t v_{ss}(t, S_t))$. Thus, (4.43) yields the PDE

$$v_t(t, s) + \frac{1}{2}\frac{\sigma^2 s^2 v_{ss}(t, s)}{1 - \lambda(y)s v_{ss}(t, s)} = 0. \tag{4.45}$$

Note that this pricing pde is (structurally) very similar to the equations derived in [LY05, FP11, BLZ17]. Due to the singularity at $\lambda(y)sv_{ss} = 1$ in (4.45), constraints on sv_{ss} (upper bound $\overline{\gamma} : \mathbb{R}_+ \to \mathbb{R}$) need to be imposed in order to have a well-posed pde.

Following the analysis in [BLZ17], it turns out that (4.45) characterizes the minimal superhedging price, after appropriate gamma constraints are imposed. Indeed, let's write

(4.36) as

$$\mathrm{d}\Theta_t = \sigma_\Theta^{a,b}(S_t)\,\mathrm{d}S_t + \mu_\Theta^{a,b}(S_t)\,\mathrm{d}t,$$
$$\text{with}\quad S_t\sigma_\Theta^{a,b}(S_t) = \frac{b_t}{\sigma + \lambda(Y_t)b_t},\quad \mu_\Theta^{a,b}(S_t) = a_t - \xi_t S_t \sigma_\Theta^{a,b}(S_t).$$

Like in [BLZ17], we restrict the admissible trading strategies $\Theta = (a, b)$ to these with Lipschitz continuous and bounded a, b, for which b is an Itô diffusion with Lipschitz continuous and bounded drift and diffusion processes, and such that $S_t\sigma_\Theta^{a,b}(S_t)$ is bounded from above by $\overline{\gamma}(S_t)$ and also bounded from below. Then the arguments in [BLZ17] would carry over to our setup, under conditions on $\overline{\gamma}$ and λ (cf. Remark 4.7.1 below), and would give that the minimal superhedging price

$$v_{\overline{\gamma}}(t, y, s) = \inf\{p \mid \exists \text{ admissible } \Theta \text{ so that (4.39) holds}\}$$

satisfies $v_{\overline{\gamma}}(t, y, s) = v_{\overline{\gamma}}^y(t, s)$, where $v_{\overline{\gamma}}^y(t, s)$ is the unique viscosity solution of the pricing pde

$$\mathbb{F}^y[\varphi](t, s) := \min\left\{-\varphi_t(t, s) - \frac{1}{2}\frac{\sigma^2 s^2 \varphi_{ss}(t, s)}{1 - \lambda(y)s\varphi_{ss}(t, s)}, \overline{\gamma}(s) - s\varphi_{ss}\right\} = 0 \quad \text{on } [0, T)\times\mathbb{R}_+, \tag{4.46}$$

with terminal condition given by the face-lifted payoff $\hat{g}$, where $\hat{g}$ is the smallest function above g that is a viscosity supersolution of the equation $\overline{\gamma} - s\varphi_{ss} \geq 0$.

We conclude this section by stressing some features of the minimal superhedging price in this case and pointing important differences to the case of non-covered options.

Remark 4.7.1. The arguments from [BLZ17] could be adapted to the present setup for bounded continuous $\overline{\gamma}$ that satisfies

$$\sup_{y\in\mathbb{R}, s\in\mathbb{R}_+} \frac{\sigma s\overline{\gamma}(s)}{1 - \lambda(y)\overline{\gamma}(s)} \in \mathbb{R}_+,$$

and continuous bounded payoffs g. The main reason is that for every $y \in \mathbb{R}$ and $s \in \mathbb{R}_+$, the map $M \in (-\infty, \gamma(s)] \mapsto \frac{\sigma^2 sM}{1-\lambda(y)M}$ is non-decreasing and convex, like in [BLZ17, Remark 3.1], ensuring that the smoothing techniques from [BLZ17, Section 3.1] go through here as well.

Remark 4.7.2.
1. The resilience function h does not appear in the pricing pde (4.46). Note that this is different from the results in Section 4.4.1, where the resilience function enters the pricing equation in a non-trivial way. However, the price of a covered option will depend on the initial level of impact y through λ.

2. The minimal superheding price is decreasing in the impact λ in the sense that if $\lambda \geq \tilde{\lambda}$, then $v_{\overline{\gamma}}^{\lambda} \geq v_{\overline{\gamma}}^{\tilde{\lambda}}$, and $v_{\overline{\gamma}}^{\lambda} \geq v^{\mathrm{BS}}$, where v^{BS} is the Black-Scholes price for the option with payoff g, i.e. v^{BS} solves $-\partial_t v^{\mathrm{BS}} - 1/2\sigma^2 s^2 \partial_{ss}^2 v^{\mathrm{BS}} = 0$ on $[0, T) \times \mathbb{R}_+$ with terminal condition $v^{\mathrm{BS}}(T, \cdot) = g(\cdot)$, see [BLZ17, Remark 2.9].

4.8 Proofs

Here we provide the proofs delegated from Section 4.3. In Section 4.8.1 we give the full details for the proof of the DPP, see equation (4.17), and after that we proceed to the proof of Theorem 4.4.9. Recall that in this case $f(x) = \exp(\lambda x)$ for $\lambda > 0$ and thus the effective price simplifies to $\mathcal{S}(s, y, \theta) = se^{-\lambda\theta} \equiv \mathcal{S}(s, \theta)$, i.e. the level of impact is not needed in order to determine the price change of a block trade, given the price before the trade. We consider strategies taking values in $\mathcal{K} = [-K, +\infty)$ for $K > 0$, yielding gradient constraints in the pde. This was needed because of the singularity of the pde by the form of the optimal strategy in (4.30).

First, we verify in Section 4.8.2 that if the pricing pde ($\mathbf{PDE}^{\delta}$) admits a sufficiently smooth classical solution, then a replicating strategy in feedback form can be constructed. Such a construction will be needed also for the contradiction argument in the proof of the subsolution property in Section 4.8.3 where, using smooth test functions, one will need to construct locally strategies which, roughly speaking, behave like replicating strategies. The viscosity property proofs are collected in Section 4.8.3 and in Section 4.8.4 we prove a comparison result that implies in particular uniqueness.

4.8.1 Proof of DPP

In this section, we outline the proof of the dynamic programming principle for the functions $\bar{w}$ and $\bar{w}_k$ defined in equation (4.17), needed in the proof of Theorem 4.3.1. The ideas are essentially contained in [ST02] but here we detail them for completeness. In what follows, we will consider $(\mathcal{F}_t)$-stopping times valued in $[0, T]$ and the set of all such stopping times will be denoted by $\mathcal{T}$.

First we will collect important properties of our setup, in particular a structural property of the set of controls Γ (or any Γ_k) and conditions that the state space process Z from (4.15) satisfies. First note that Γ is stable under concatenation, meaning that for all $\gamma_1,\ \gamma_2 \in \Gamma$ and stopping times τ, the τ-concatenation of (γ_1, γ_2), given by $\gamma_1 \oplus_\tau \gamma_2 := \gamma_1 \mathbb{1}_{[\![0,\tau[\![} + \gamma_2 \mathbb{1}_{[\![\tau,T]\!]}$, is still an element of Γ. It is also clear that if $\gamma_1 \in \Gamma_{k_1}$ and $\gamma_2 \in \Gamma_{k_2}$, then $\gamma_1 \oplus_\tau \gamma_2 \in \Gamma_{k_1+k_2}$. Since $\mathcal{F}$ is countably generated, we can also endow Γ with a suitable topology as in [BLZ16, Appendix] that turns Γ into a complete and separable metric space, see also [ST02, Section 2.5]. In particular, the "stability under measurable selection" condition A2 from [ST02] holds, cf. [ST02, Lemma 2.1]: for any stopping time τ and any measurable map $\phi : (\Omega, \mathcal{F}_\tau) \to (\Gamma, \mathcal{B}(\Gamma))$, there exists $\gamma \in \Gamma$ such that

$$\phi = \gamma \quad \text{on } [\![\tau, T]\!] \times \Omega. \quad \text{Leb} \otimes \mathbb{P} \text{ a.e.}$$

We now consider the state process Z as a function of (random) initial conditions, i.e. $Z : \mathcal{S} \to \mathbb{H}^0(\mathbb{R}^4)$, where $\mathcal{S}$ is the set of all pairs (τ, ζ) with $\tau \in \mathcal{T}$ and $\mathbb{R}_+ \times \mathbb{R}^3$-valued $\mathcal{F}_\tau$-measurable and square-integrable random variables ζ, and $\mathbb{H}^0(\mathbb{R}^4)$ is the set of all progressively measurable càdlàg processes $X : [0, T] \times \Omega \to \mathbb{R}^4$. As such, we list the following properties that the state process in our setup satisfies.

Z1 *Initial data:* $Z_{\tau-}^{\tau,\zeta,\gamma} = \zeta$.

Z2 *Consistency with deterministic initial data:* for all $(t, z) \in [0, T] \times \mathbb{R}_+ \times \mathbb{R}^3$ and any bounded measurable function f

$$\mathbb{E}[f(Z_s^{\tau,\zeta,\gamma}) \mid (\tau, \zeta) = (t, z)] = \mathbb{E}[f(Z_s^{t,z,\gamma})] \quad \forall s \in [t, T].$$

Z3 *Pathwise uniqueness:* for all $\tau_1, \tau_2 \in \mathcal{T}$ with $\tau_1 \le \tau_2$, we have

$$Z^{\tau_1,\zeta,\gamma} = Z^{\tau_2,\xi,\gamma} \text{ on } [\![\tau_2, T]\!], \text{ where } \xi := Z_{\tau_2}^{\tau_1,\zeta,\gamma}.$$

Z4 *Causality:* if two admissible controls γ_1 and γ_2 are equal between two stopping times $\tau_1 \le \tau_2 \in \mathcal{T}$, i.e. $\gamma_1 = \gamma_2$ on $[\![\tau_1, \tau_2]\!]$, then

$$Z^{\tau_1,\zeta,\gamma_1} = Z^{\tau_1,\zeta,\gamma_2} \text{ on } [\![\tau_1, \tau_2]\!].$$

Z5 *Measurability:* the map

$$(t, z, \gamma) \in [0, T] \times (\mathbb{R}_+ \times \mathbb{R}^3) \times \Gamma \mapsto Z_T^{t,z,\gamma} \in \mathbb{R}^4$$

is Borel measurable.

That our state process Z satisfies **Z1**–**Z5** follows as in [ST02, Proof of Prop. 6.1]; see also [BLZ16, Appendix] for Z5.

Now we are ready to provide in details the arguments needed in the proof of Theorem 4.3.1. For the proof of **Part 1**, we were in the following situation: there exists $\theta \in \mathcal{K}$ and some $\gamma \in \mathcal{G}(t, z)$ for $z = (\mathcal{S}(s, y, -\theta), y + \theta, \theta, v)$.

Let $\tau \ge t$ be a stopping time in $\mathcal{T}$. The pathwise uniqueness property **Z3** yields

$$Z_T^{t,z,\gamma} = Z_T^{\tau, Z_\tau^{t,z,\gamma}, \gamma} \in \mathfrak{G} \text{ a.s.}$$

Let $\mu = \mu_{\tau,t,z}$ be the law of $(\tau, Z_\tau^{t,z,\gamma})$ on $[0, T] \times \mathbb{R}^4$. We show that $\gamma \in \mathcal{G}(t', z')$ for μ-almost all (t', z'). Indeed, by **Z2** we have for every (t', z')

$$\mathbb{P}\left(Z_T^{t',z',\gamma} \in \mathfrak{G}\right) = \mathbb{P}\left(Z_T^{\tau, Z_\tau^{t,z,\gamma}, \gamma} \in \mathfrak{G} \mid (\tau, Z_\tau^{t,z,\gamma}) = (t', z')\right).$$

Hence,

$$\begin{aligned} \int \mathbb{P}\left(Z_T^{t',z',\gamma} \in \mathfrak{G}\right) \mathrm{d}\mu &= \mathbb{E}\left(\mathbb{P}\left(Z_T^{\tau, Z_\tau^{t,z,\gamma}, \gamma} \in \mathfrak{G} \mid (\tau, Z_\tau^{t,z,\gamma}) = (t', z')\right)\right) \\ &= \mathbb{P}\left(Z_T^{\tau, Z_\tau^{t,z,\gamma}, \gamma} \in \mathfrak{G}\right) \\ &= \mathbb{P}\left(Z_T^{t,z,\gamma} \in \mathfrak{G}\right) = 1. \end{aligned}$$

Therefore for μ-almost all (t', z'), $\mathbb{P}\left(Z_T^{t',z',\gamma} \in \mathfrak{G}\right) = 1$ and hence $\gamma \in \mathcal{G}(t', z')$ also.

Therefore, we get $\gamma \in \mathcal{G}(\tau, Z_\tau^{t,z,\gamma})$. This means by definition of $\bar{w}$ (cf. (4.17)) that

$$V_\tau^{\mathrm{liq},t,z,\gamma} \geq \bar{w}(\tau, S_\tau^{t,z,\gamma}, Y_\tau^{t,z,\gamma}, \Theta_\tau^{t,z,\gamma}),$$

and this is what was needed for the proof of Part 1 in Theorem 4.3.1.

Now we proceed with the arguments needed for **Part 2**. In this case, we have for some $\theta \in \mathcal{K} \cap [-k, k]$, control $\gamma \in \Gamma_k$ and a stopping time $\tau \in [t, T]$,

$$V_\tau^{\mathrm{liq},t,z,\gamma} > \bar{w}_{k+1}(S_\tau^{t,z,\gamma}, Y_\tau^{t,z,\gamma}, \Theta_\tau^{t,z,\gamma}), \tag{4.47}$$

where $z := (\mathcal{S}(s, y, -\theta), y + \theta, \theta, v)$, and we need to prove that $v \geq \bar{w}_{2k+1}(t, z)$.

For this purpose, let again μ be the law of $(\tau, Z_\tau^{t,z,\gamma})$ on $[0, T] \times \mathbb{R}^4$. Then [ST02, Lemma 3.1] gives the existence of a Borel measurable function $\phi_\mu : (D, \mathcal{B}(D)) \to (\Gamma_k, \mathcal{B}(\Gamma_{k+1}))$ such that

$$\phi_\mu(t, z) \in \mathcal{G}_{k+1}(t, z) \text{ for } \mu - \text{a.a. } (t, z) \in [0, T] \times \mathbb{R}^4,$$

where $D := \{(t, z) \in [0, T] \times \mathbb{R}^4 \mid \mathcal{G}(t, z) \neq \emptyset\}$. Because of (4.47) we clearly have $(\tau, Z_\tau^{t,z,\gamma}) \in D$ a.s. Now, using the stability under measurable selection property of our control set Γ_{k+1} and the measurable selector ϕ_μ, we get the existence of a control $\gamma_1 \in \Gamma_{k+1}$ such that for all bounded and measurable functions f

$$\mathbb{E}\left(f(Z_T^{t',z',\gamma_1})\right) = \mathbb{E}\left(f(Z_T^{\tau, Z_\tau^{t,z,\gamma}, \phi_\mu(\tau, Z_\tau^{t,z,\gamma})}) \big| (\tau, Z_\tau^{t,z,\gamma}) = (t', z').\right) \tag{4.48}$$

Set now $\hat{\gamma} = \gamma \oplus_\tau \gamma_1$. Then $\hat{\gamma} \in \Gamma_{2k+1}$, and we have

$$\begin{aligned}
Z_T^{t,z,\hat{\gamma}} &= Z_T^{\tau, Z_\tau^{t,z,\hat{\gamma}}, \hat{\gamma}} && \text{by } \mathbf{Z3} \\
&= Z_T^{\tau, Z_\tau^{t,z,\gamma}, \hat{\gamma}} && \text{by } \mathbf{Z4} \text{ since } \hat{\gamma} = \gamma \text{ on } [\![t, \tau]\!] \\
&= Z_T^{\tau, Z_\tau^{t,z,\gamma}, \gamma_1} && \text{by } \mathbf{Z4} \text{ since } \hat{\gamma} = \gamma_1 \text{ on } [\![\tau, T]\!] \\
&\in \mathfrak{G} \text{ a.s.} && \text{by (4.48) for } f = \mathbb{1}_{\mathfrak{G}}.
\end{aligned}$$

In particular, $\hat{\gamma} \in \mathcal{G}_{2k+1}(t, z)$ and hence $v \geq \bar{w}_{2k+1}(t, z)$, that was to be proved.

4.8.2 Verification argument for exponential impact function

Suppose that $w \in C^{1,2,1}([0, T] \times \mathbb{R}_+ \times \mathbb{R})$ is such that for every $(t, s, y) \in [0, T] \times \mathbb{R}_+ \times \mathbb{R}$

1. $\theta[w](t, s, y) \in \mathcal{K}$,

2. $\mathcal{L}^{\theta[w](t,s,y)} w(t, s, y) = 0$ when $t < T$, and

3. $w(T, s, y) = H(s, y)$.

Suppose also that w is sufficiently regular so that there exists an admissible strategy $\Theta \in \Gamma$ of the form

$$\begin{aligned} &\Theta_t = 1/\lambda \log(\lambda w_S(t, \mathcal{S}(S_t, \Theta_t), Y_t - \Theta_t) + 1) \quad \text{for } t \in [0, T), \\ &\Theta_T = 0, \qquad \text{i.e. } \Delta\Theta_T = \Theta_{T-}. \end{aligned} \tag{4.49}$$

In particular, $\Theta_0 = 1/\lambda \log(\lambda w_S(0, s, y) + 1)$ and $\Delta\Theta_T \in \mathcal{K}$. Consider the self-financing portfolio (β, Θ) with $\beta_{0-} = w(0, s, y)$. Then as in Remark 4.4.8 we get

$$V_T^{\text{liq}}(\Theta) = H(S_T, Y_T^\Theta), \qquad \Theta_T = 0.$$

By definition of H, we will have enough capital to (super-)replicate the European claim with payoff (g_0, g_1) with a possible additional jump trade (provided that the infima in the definition of H, cf. Lemma 4.4.1, are attained). Hence, (β, Θ) will be a (super-)replicating strategy for the European claim (g_0, g_1) with initial capital $w(0, s, y)$, meaning that its price is exactly $w(0, s, y)$.

Remark 4.8.1 (On the form of a replicating strategy)**.** To construct a replicating strategy (4.49), similarly to Remark 4.4.8 we applying Itô's formula to get for $t < T$

$$\begin{aligned} \mathrm{d}\Theta_t &= \frac{1}{\lambda}\left(\frac{1}{\lambda w_S + 1}\,\mathrm{d}(\lambda w_S + 1) - \frac{1}{2(\lambda w_S + 1)^2}\,\mathrm{d}[\lambda w_S + 1]_t\right) \\ &= a(t, \mathcal{S}_t, \mathcal{Y}_t^\Theta, \Theta_t)\,\mathrm{d}t + b(t, \mathcal{S}_t, \mathcal{Y}_t^\Theta)\,\mathrm{d}W_t, \end{aligned}$$

where for $\mathcal{S}_t := \mathcal{S}(S_t, \Theta_t)$ and $\mathcal{Y}_t^\Theta = Y_t^\Theta - \Theta_t$ we set

$$\begin{aligned} a(t, \mathcal{S}_t, \mathcal{Y}_t^\Theta, \Theta_t) &:= \frac{1}{\lambda w_S + 1}\Big(w_{tS} + w_{SS}\mathcal{S}_t(\mu_t - \lambda h(Y_t^\Theta)) - w_{SY} h(Y_t^\Theta) + \\ &\qquad\qquad + \frac{1}{2} w_{SSS}\sigma^2 \mathcal{S}_t^2 - \frac{\lambda^2\sigma^2\mathcal{S}_t^2 w_{SS}}{2(\lambda w_S + 1)}\Big), \\ b(t, \mathcal{S}_t, \mathcal{Y}_t^\Theta) &:= \frac{\sigma\mathcal{S}_t w_{SS}}{\lambda w_S + 1}; \end{aligned}$$

all the derivatives of w above are evaluated at $(t, \mathcal{S}(S_t, \Theta_t), Y_t - \Theta_t)$. Thus, a replicating strategy gives a solution of the following system for $t \in [0, T)$

$$\begin{aligned} \mathrm{d}\mathcal{S}_t &= \mathcal{S}_t[(\mu_t - \lambda h(Y_t))\,\mathrm{d}t + \sigma\,\mathrm{d}W_t], \\ \mathrm{d}\Theta_t &= a(t, \mathcal{S}_t, \mathcal{Y}_t^\Theta, \Theta_t)\,\mathrm{d}t + b(t, \mathcal{S}_t, \mathcal{Y}_t^\Theta)\,\mathrm{d}W_t, \\ \mathrm{d}\mathcal{Y}_t^\Theta &= -h(\mathcal{Y}_t^\Theta + \Theta_t)\,\mathrm{d}t, \end{aligned} \tag{4.50}$$

with initial condition $\mathcal{S}_0 = s$, $\mathcal{Y}_0^\Theta = y$ and $\Theta_0 = 1/\lambda \log(\lambda w_S(0, s, y) + 1)$. Under certain conditions (e.g. on H), one could derive that the coefficients a and b are (locally) bounded Lipschitz continuous functions, thus giving an admissible replicating strategy Θ.

4.8.3 Viscosity property for exponential impact function

Now we prove the viscosity property from Section 4.4.2.

Theorem 4.8.2. *The function w_* from* (4.24) *is a viscosity supersolution of* ($\mathbf{PDE}^\delta$) *on* $[0,T) \times \mathbb{R}_+ \times \mathbb{R}$ *with the boundary condition* ($\mathbf{BC}^\delta$) *on* $\{T\} \times \mathbb{R}_+ \times \mathbb{R}$.

Proof. First, let $(t_0, s_0, y_0) \in [0,T) \times \mathbb{R}_+ \times \mathbb{R}$ and $\varphi \in C_b^\infty([0,T] \times \mathbb{R}_+ \times \mathbb{R})$ be a smooth function such that

$$\text{(strict)} \quad \min_{[0,T]\times\mathbb{R}_+\times\mathbb{R}} (w_* - \varphi) = (w_* - \varphi)(t_0, s_0, y_0) = 0.$$

Case 1: Suppose that $\mathcal{H}_\mathcal{K}\varphi(t_0, s_0, y_0) < 0$. By continuity of the operator $\mathcal{H}_\mathcal{K}$ there exists a neighborhood $\mathcal{O} \subset [0,T] \times \mathbb{R}_+ \times \mathbb{R}$ of (t_0, s_0, y_0) such that $\mathcal{H}_\mathcal{K}\varphi(t,s,y) < -\varepsilon$ in $\mathcal{O}$ for some $\varepsilon > 0$. Therefore, after possibly decreasing the neighbourhood $\mathcal{O}$, there exists a constant $k_\varepsilon > 0$ such that

$$s|\varphi_S(t,s,y) + 1/\lambda - e^{\lambda\theta}/\lambda| \geq k_\varepsilon \qquad \forall \theta \in \mathcal{K}, \forall (t,s,y) \in \mathcal{O}. \tag{4.51}$$

Let $(t_n, s_n, y_n)_n \subset \mathcal{O}$ be a sequence converging to (t_0, s_0, y_0) with $w(t_n, s_n, y_n) \to$ $\to w_*(t_0, s_0, y_0)$ (note that w_* is the lower-semicontinuous envelope of w), cf. Lemma 5.3.9. Set $v_n := w(t_n, s_n, y_n) + 1/n$. Since $v_n > w(t_n, s_n, y_n)$, Theorem 4.3.1 implies the existence of $\theta_n \in \mathcal{K}$ and strategies $\gamma_n \in \Gamma$ such that for stopping times τ_n (to be chosen later) we have $\mathbb{P}$-a.s.

$$V^{\mathrm{liq}, t_n, z_n, \gamma_n}_{t\wedge\tau_n} \geq w(\cdot, \mathcal{S}(S^{t_n,z_n,\gamma_n}, \Theta^{t_n,z_n,\gamma_n}), Y^{t_n,z_n,\gamma_n} - \Theta^{t_n,z_n,\gamma_n})_{t\wedge\tau_n}, \tag{4.52}$$

where $z_n = (s_n e^{\lambda\theta_n}, y_n + \theta_n, \theta_n, v_n)$. For notational convenience in what follows we will use superscript n instead of superscript (t_n, z_n, γ_n) and $\mathcal{S}^n := \mathcal{S}(S^{t_n,z_n,\gamma_n}, \Theta^{t_n,z_n,\gamma_n})$, $\mathcal{Y}^n := Y^{t_n,z_n,\gamma_n} - \Theta^{t_n,z_n,\gamma_n}$.

Take $\tau_n = \inf\{t \geq t_n \;\; (t, \mathcal{S}^n, \mathcal{Y}^n_t) \in \partial_p\mathcal{O}\}$, where $\partial_p\mathcal{O}$ denotes the parabolic boundary of the open region $\mathcal{O}$. In particular, $\tau_n \leq T$. Since $w \geq w_* \geq \varphi$ and $w_* - \varphi$ has a strict local minimum at (t_0, s_0, y_0), there exists $\iota > 0$ such that

$$(w - \varphi)(\tau_n, \mathcal{S}^n_{\tau_n}, \mathcal{Y}^n_{\tau_n}) \geq \iota.$$

Hence, $V^{\mathrm{liq},n}_{\tau_n} - \varphi(\tau_n, \mathcal{S}^n_{\tau_n}, \mathcal{Y}^n_{\tau_n}) \geq \iota$. Now, Lemma 4.3.3 together with the fact that $\mathcal{S}^n_{t_n} = s_n$, $\mathcal{Y}^n_{t_n} = y_n$, gives that $\mathbb{P}$-a.s.

$$\begin{aligned} \iota \leq & v_n - \varphi(t_n, s_n, y_n) + \\ & - \int_{t_n}^{\tau_n} \mathcal{S}^n_u(\varphi_S(\mathcal{S}^n_u, \mathcal{Y}^n_u) + 1/\lambda - e^{\lambda\Theta^n_u}/\lambda)\,(\sigma\,\mathrm{d}W_u + \zeta^n_u\,\mathrm{d}u)\,, \end{aligned} \tag{4.53}$$

where

$$\zeta^n_t := \eta^n_t - \frac{\mathcal{L}^{\Theta^n_t}\varphi}{\mathcal{S}^n_t(\varphi_S(\mathcal{S}^n_t, \mathcal{Y}^n_t) + 1/\lambda - e^{\lambda\Theta^n_t}/\lambda)} \quad \text{for } t \in [t_n, \tau_n]$$

with $\eta_t^n := \mu_t - \lambda h(Y_t^n)$. Note that ζ_t^n is well-defined on $[t_n, \tau_n]$ and uniformly bounded because of (4.51) and the fact that Y^n is bounded since Θ^n is so. Hence, Girsanov's theorem gives a measure $\mathbb{P}^n$ which is equivalent to $\mathbb{P}$ on $[t_n, \tau_n]$ such that

$$\int_{t_n}^{t\wedge\tau_n} \mathcal{S}_u(\varphi_S(\mathcal{S}_u^n, \mathcal{Y}_u^n) + 1/\lambda - e^{\lambda\Theta_u}/\lambda)\,(\sigma\,\mathrm{d}W_u + \zeta_u^n\,\mathrm{d}u)$$

is a square-integrable martingale under $\mathbb{P}^n$ on every compact time-interval (the integrand of the stochastic integral is uniformly bounded because of the definition of τ_n, the continuity of φ_S and the boundedness of the range of Θ). Hence, taking expectation under $\mathbb{P}^n$ of the right-hand side of (4.53) leads to

$$v_n - \varphi(t_n, s_n, y_n) \geq \iota > 0,$$

which is a contradiction since by our choice of v_n and the sequence $(t_n, s_n, y_n)_n$

$$v_n - \varphi(t_n, s_n, y_n) \longrightarrow w_*(t_0, s_0, y_0) - \varphi(t_0, s_0, y_0) = 0.$$

Case 2: From Case 1 we know that $\mathcal{H}_{\mathcal{K}}\varphi(t_0, s_0, y_0) \geq 0$. Hence

$$\theta[\varphi](t_0, s_0, y_0) = 1/\lambda \log(\lambda\varphi_S(t_0, s_0, y_0) + 1)$$

is well-defined (in a neighborhood of (t_0, s_0, y_0)) and suppose that $\mathcal{L}^{\theta[\varphi]}\varphi(t_0, s_0, y_0) < 0$. By continuity of the operator $\mathcal{L}$, there exits an open neighborhood $\mathcal{O} \subset [0, T] \times \mathbb{R}_+ \times \mathbb{R}$ of (t_0, s_0, y_0) and some $r, \varepsilon > 0$ such that

$$\mathcal{L}^\theta\varphi(t, s, y) < -\varepsilon \quad \forall (t, s, y) \in \mathcal{O},\ \forall \theta \in (\theta[\varphi](t, s, y) - r\ ,\ \theta[\varphi](t, s, y) + r).$$

In particular, by continuity of the functions involved we have (after possibly decreasing the open set $\mathcal{O}$) that for every $(t, s, y) \in \mathcal{O}$ and for some $r' > 0$

$$\mathcal{L}^\theta\varphi(t, s, y) < -\varepsilon \quad \text{whenever} \quad |\varphi_S(t, s, y) + 1/\lambda - e^{\lambda\theta}/\lambda| \leq r'.$$

As in Case 1, consider a sequence (t_n, s_n, y_n) in $\mathcal{O}$ which converges to (t_0, s_0, y_0) and such that $w(t_n, s_n, y_n) \to w_*(t_0, s_0, y_0)$. Set $v_n := w(t_n, s_n, y_n) + 1/n$ and let $\theta_n \in \mathcal{K}$ and strategies $\gamma_n \in \Gamma$ be such that the dynamic programming principle (4.52) holds for the stopping times τ_n that are the first exit times of $(\cdot, \mathcal{S}^n, \mathcal{Y}^n)$ from the set $\mathcal{O}$. Now, the proof follows the same lines as in Case 1 with the following adjustment:

$$\begin{aligned}
V^{\mathrm{liq},n}_{t\wedge\tau_n} - \varphi(\cdot,\mathcal{S}^n,\mathcal{Y}^n)_{t\wedge\tau_n} &= v_n - \varphi(t_n,s_n,y_n) \\
&- \int_{t_n}^{t\wedge\tau_n} \mathcal{S}^n_u(\varphi_S + 1/\lambda - e^{\lambda\Theta^n_u}/\lambda)\,(\sigma\,\mathrm{d}W_u + \zeta^n_u\,\mathrm{d}u) + \\
&+ \int_{t_n}^{t\wedge\tau_n} \mathcal{L}^{\Theta^n_u}\varphi(\mathcal{S}^n_u,\mathcal{Y}^n_u)\mathbb{1}_{\{|\varphi_S+1/\lambda-e^{\lambda\Theta^n_u}/\lambda|\le r'\}}\,\mathrm{d}u \\
&\le v_n - \varphi(t_n,s_n,y_n) - \int_{t_n}^{t\wedge\tau_n} \mathcal{S}^n_u(\varphi_S + 1/\lambda - e^{\lambda\Theta^n_u}/\lambda)\,(\sigma\,\mathrm{d}W_u + \zeta^n_u\,\mathrm{d}u)\,,
\end{aligned}$$

where for $t \in [t_n,\tau_n]$

$$\zeta^n_t := \eta^n_t - \frac{\mathcal{L}^{\Theta^n_t}\varphi}{\mathcal{S}^n_t(\varphi_S + 1/\lambda - e^{\lambda\Theta^n_t}/\lambda)}\mathbb{1}_{\{|\varphi_S+1/\lambda-e^{\lambda\Theta^n_t}/\lambda|\ge r'\}}$$

(the functions φ and φ_S in the expressions above are evaluated at $(\mathcal{S}^n,\mathcal{Y}^n)$). The contradiction now follows after taking expectation under $\mathbb{P}^n \approx \mathbb{P}$ on $[t_n,\tau_n]$ and letting $n\to\infty$.

Boundary condition. Let $(s_0,y_0)\in\mathbb{R}_+\times\mathbb{R}$ and φ be a smooth function such that

$$\text{(strict)}\quad \min_{[0,T]\times\mathbb{R}_+\times\mathbb{R}}(w_* - \varphi) = (w_* - \varphi)(T,s_0,y_0) = 0.$$

Suppose that

$$\min\{w_*(T,s_0,y_0) - H(s_0,y_0), \mathcal{H}_{\mathcal{K}}\varphi(T,s_0,y_0)\} < 0.$$

The case $\mathcal{H}_{\mathcal{K}}\varphi(T,s_0,y_0) < 0$ leads to a contradiction by the same arguments as in Case 1 above, using that $\mathcal{H}_{\mathcal{K}}\varphi < 0$ is a small neighborhood of (T,s_0,y_0). Hence we have $\mathcal{H}_{\mathcal{K}}\varphi(T,s_0,y_0) \ge 0$.

Now, if $w_*(T,s_0,y_0) < H(s_0,y_0)$ then also $\varphi(T,s_0,y_0) - H(s_0,y_0) < 0$. After possibly modifying the test function φ by $(t,s,y)\mapsto\varphi(t,s,y) - \sqrt{T-t}$, we can assume that $\partial_t\varphi(t,s,y)\to+\infty$ when $t\to T$, uniformly on compacts. Hence, in an ε-neighborhood $[T-\varepsilon,T)\times B_\varepsilon(s_0,y_0)$ around (T,s_0,y_0) we have $\mathcal{L}^{\theta[\varphi]}\varphi < 0$. Moreover, after possibly decreasing ε we have $\varphi(T,\cdot) \le H(\cdot) - \iota_1$ on $B_\varepsilon(s_0,y_0)$ for some $\iota_1 > 0$. We argue as in Case 1 and 2 above (by starting from (t_n,s_n,y_n) in $[T-\varepsilon,T)\times B_\varepsilon(s_0,y_0)$, with $(t_n,s_n,y_n)\to(T,s_0,y_0)$ and $w(t_n,s_n,y_n)\to w_*(T,s_0,y_0)$, stopping at the (parabolic) boundary at a time τ_n, and using $w(T,\cdot) = H(\cdot)$) to get

$$V^{\mathrm{liq},n}_{\tau_n} - \varphi(\cdot,\mathcal{S}(S^n,\Theta^n),Y^n - \Theta^n)_{\tau_n} \ge \iota_1\wedge\iota_2,$$

where $\iota_2 := \inf_{[T-\varepsilon,T)\times\partial B_\varepsilon(s_0,y_0)}(w_* - \varphi) > 0$, and a contradiction follows as in Case 2 above.

□

Now we prove the subsolution property.

Theorem 4.8.3. *The function w^* from* (4.25) *is a viscosity subsolution of* ($\mathbf{PDE}^\delta$) *on* $[0,T) \times \mathbb{R}_+ \times \mathbb{R}$ *with the boundary condition* ($\mathbf{BC}^\delta$) *on* $\{T\} \times \mathbb{R}_+ \times \mathbb{R}$.

Proof. The proof is very similar to proof of the subsolution property in [BLZ16, Theorem 3.7]. The reason is that in this case, the gradient constraints will ensure that a test function φ, that would possibly contradict the subsolution property, should satisfy $\mathcal{H}_\mathcal{K}\varphi > 0$ locally and hence would be sufficiently "nice" to define (locally) control processes (employing the verification argument in Remark 4.8.1) that would lead to a contradiction like in [BLZ16]. For completeness, we outline differences and sketch the main steps.

Let φ be a $C_b^\infty([0,T], \mathbb{R}_+ \times \mathbb{R})$ test function such that $(t_0, s_0, y_0) \in [0,T] \times \mathbb{R}_+ \times \mathbb{R}$ is a strict (local) maximum of $w^* - \varphi$, i.e.

$$\text{(strict)} \max_{[0,T]\times\mathbb{R}_+\times\mathbb{R}} (w^* - \varphi) = (w^* - \varphi)(t_0, s_0, y_0) = 0.$$

First assume that $t_0 < T$. To ease the notations, we will use the variable x to denote the pair (s, y). Because of the special form of the DPP, Part 2, cf. Theorem 4.3.1, we need to employ w_k (instead of w as we did in the proof of the supersolution property). By Lemma 5.3.12 there exists a sequence $(k_n, t_n, x_n)_{n\geq 1}$ such that $k_n \to \infty$, (t_n, x_n) is a local maxima of $w_{k_n}^* - \varphi$, and $(t_n, x_n, w_{k_n}(t_n, x_n)) \to (t_0, x_0, w^*(t_0, x_0))$.

Assume that $\mathcal{F}_\mathcal{K}[\varphi](t_0, x_0) > 0$ and let $\varphi_n(t,x) = \varphi(t,x) + |t-t_n|^2 + |y-y_n|^2 + |s-s_n|^4$. Then $\mathcal{F}_\mathcal{K}[\varphi_n] > 0$ in a neighborhood B of (t_0, x_0) that contains (t_n, x_n), for all n large enough. Since we will be working on the local neighborhood B where also $\mathcal{H}_\mathcal{K}\varphi_n > 0$, we can modify (in a smooth way) the functions h and φ_n outside of B to be supported on a slightly bigger compact set. Thus, (after possibly changing $n \geq 1$) there exists $\gamma_n \in \Gamma_{k_n}$ such that

$$\Theta_t^{t_n,z_n,\gamma_n} = 1/\lambda \cdot \log\left(\lambda \frac{\partial \varphi_n}{\partial s}(t, \mathcal{S}_t^{t_n,z_n,\gamma_n}, \mathcal{Y}_t^{t_n,z_n,\gamma_n}) + 1\right), \quad t \geq t_n,$$

where for $z_n = (s_n, y_n, 0, w_{k_n}(t_n, x_n) - n^{-1})$ we set $\mathcal{S}_t^{t_n,z_n,\gamma_n} = \mathcal{S}(S_t^{t_n,z_n,\gamma_n}, \Theta_t^{t_n,z_n,\gamma_n})$ and $\mathcal{Y}_t^{t_n,z_n,\gamma_n} = (Y-\Theta)_t^{t_n,z_n,\gamma_n}$, see Remark 4.8.1. Let τ_n be the first time after t_n at which the process $(\mathcal{S}_t^{t_n,z_n,\gamma_n}, \mathcal{Y}_t^{t_n,z_n,\gamma_n})_{t\geq t_n}$ leaves B. Like in [BLZ16, Proof of Thm. 3.7] we conclude, applying Itô's formula, using Lemma 4.3.3 and $F[\varphi_n] > 0$ on B, that $\mathbb{P}$-a.s.

$$V_{\tau_n}^{\text{liq},t_n,z_n,\gamma_n} \geq \varphi_n(\tau_n, \mathcal{S}_{\tau_n}^{t_n,z_n,\gamma_n}, (Y-\Theta)_{\tau_n}^{t_n,z_n,\gamma_n}) + v_n - \varphi_n(t_n, x_n).$$

Now, a contradiction follows as in [BLZ16, Proof of Thm. 3.7, subsol. property, a]. The same arguments will be explained in details in the proof of Proposition 5.3.11, Case 1.

For the boundary condition, i.e. the case $t_0 = T$, the arguments are exactly the same as in [BLZ16, Proof of Thm. 3.7, subsol. property, b] and will be detailed for a related setup in the proof of Proposition 5.3.11, Case 2. $\square$

4.8.4 Comparison results

First we provide a comparison result for the pricing pde (**PDE**), needed for the proof of Theorem 4.4.5. Note that (**PDE**) has the following structure

$$0 = -\partial_t\varphi - \frac{\sigma^2 s^2}{2}\partial_{ss}\varphi - B_1(y, f(y)\partial_s\varphi)\partial_y\varphi - sB_2(y, f(y)\partial_s\varphi)\partial_s\varphi - sB_3(y, f(y)\partial_s\varphi), \tag{4.54}$$

where for $i = 1, 2, 3$, the functions $B_i : \mathbb{R}^2 \to \mathbb{R}$ are bounded and Lipschitz continuous. We can transform it the following way.

Lemma 4.8.4. *Let u be viscosity subsolution (resp. supersolution) of* (4.54)*. Fix $\kappa > 0$. Then $\tilde{u}$ defined by*

$$\tilde{u}(t, s, y) = e^{\kappa t}u(t, sf(y), y), \quad \forall (t, s, y) \in [0, T] \times \mathbb{R}_+ \times \mathbb{R}$$

is subsolution (resp. supersolution) of

$$\begin{aligned} 0 = \kappa\varphi - \partial_t\varphi - \tfrac{\sigma^2 s^2}{2}\partial_{ss}\varphi - B_1(y, e^{-\kappa t}\partial_s\varphi)\partial_y\varphi + \lambda(y)B_1(y, e^{-\kappa t}\partial_s\varphi)\partial_s\varphi \\ - sB_2(y, e^{-\kappa t}\partial_s\varphi)\partial_s\varphi - e^{\kappa t}sf(y)B_3(y, e^{-\kappa t}\partial_s\varphi). \end{aligned} \tag{4.55}$$

Proof. We have formally (if derivatives exist)

$$\begin{aligned} \tilde{u}_s(t, s, y) &= e^{\kappa t}f(y)u_s(t, sf(y), y) \\ \tilde{u}_{ss}(t, s, y) &= e^{\kappa t}f^2(y)u_{ss}(t, sf(y), y) \\ \tilde{u}_y(t, s, y) &= e^{\kappa t}\lambda(y)f(y)u_s(t, sf(y), y) + e^{\kappa t}u_y(t, sf(y), y) \\ &= \lambda(y)\tilde{u}_s(t, s, y) + e^{\kappa t}u_y(t, sf(y), y) \\ \tilde{u}_t(t, s, y) &= e^{\kappa t}u_t(t, sf(y), y) + \kappa e^{\kappa t}u(t, sf(y), y). \end{aligned}$$

Writing now the equation (4.54) for u at $(t, sf(y), y)$, we can read off the equation (4.55) for $\tilde{u}$. Now clearly the viscosity property of u implies the viscosity property of $\tilde{u}$ by definition of viscosity solutions. $\square$

By Lemma 4.8.4 it is now enough to prove comparison for (4.55) since this would imply a comparison result for (4.54). This is done in the following.

Theorem 4.8.5. *Let u (resp. v) be a bounded upper-semicontinuous subsolution (resp. lower-semicontinuous supersolution) on $[0, T) \times \mathbb{R}_+ \times \mathbb{R}$ of* (4.55)*. Suppose that $u \le v$ on $\{T\} \times \mathbb{R}_+ \times \mathbb{R}$. Then $u \le v$ on $[0, T] \times \mathbb{R}_+ \times \mathbb{R}$.*

Proof. Suppose by contradiction that

$$\sup_{(t,s,y)\in[0,T]\times\mathbb{R}_+\times\mathbb{R}} (u - v)(t, s, y) > 0.$$

Then we can find $R > 1$ such that

$$\sup_{(t,s,y)\in[0,T]\times\mathcal{O}_R\times[-R,R]} (u-v)(t,s,y) > 0,$$

where $\mathcal{O}_R := (1/R, R)$. In particular, there exists $\delta > 0$ and $(t_0, s_0, y_0) \in \overline{\mathcal{O}}_R \times [-R, R]$ such that $(u-v)(t_0, s_0, y_0) = \delta > 0$.

Now consider the bounded upper-semicontinuous function

$$\Phi_n(t, s_1, s_2, y_1, y_2) := u(t, s_1, y_1) - v(t, s_2, y_2) - \frac{n}{2}(s_1 - s_2)^2 - \frac{n}{2}(y_1 - y_2)^2.$$

It attains its maximum at some $(t^n, s_1^n, s_2^n, y_1^n, y_2^n) \in [0, T] \times \overline{\mathcal{O}}_R^2 \times [-R, R]^2$ by compactness of the set, and we clearly have

$$\Phi_n(t^n, s_1^n, s_2^n, y_1^n, y_2^n) \geq \delta \quad \forall n \in \mathbb{N}. \tag{4.56}$$

By the arguments in the proof of [BLZ16, Lemma 3.11] we have (after possibly passing to a subsequence)

$$n(s_1^n - s_2^n)^2 + n(y_1^n - y_2^n)^2 \to 0 \ \text{ as } n \to \infty. \tag{4.57}$$

Note that (4.57) also implies $n(s_1^n - s_2^n)(y_1^n - y_2^n) \to 0$ as $n \to \infty$.

An application of Ishii's lemma, as in [CIL92, Theorem 8.3], gives the existence of $(b^n, X^n, Y^n) \in \mathbb{R} \times S_2 \times S_2$, such that with $p^n = n(s_1^n - s_2^n)$ and $q^n = n(y_1^n - y_2^n)$

$$\begin{aligned}
&(b^n, (p^n, q^n), X^n) \in \bar{\mathcal{P}}^+_{\mathcal{O}_a} u(t^n, s_1^n, y_1^n), \\
&(b^n, (p^n, q^n), Y^n) \in \bar{\mathcal{P}}^-_{\mathcal{O}_a} v(t^n, s_2^n, y_2^n),
\end{aligned}$$

where X^n and Y^n satisfy

$$\begin{pmatrix} X^n & 0 \\ 0 & -Y^n \end{pmatrix} \leq 3n \begin{pmatrix} I_2 & -I_2 \\ -I_2 & I_2 \end{pmatrix}; \tag{4.58}$$

here S_2 denotes the set of 2×2 symmetric non-negative matrices and $I_2 \in S_2$ the identity matrix. Using the viscosity property of u and v at (t^n, s_1^n, y_1^n) and (t^n, s_2^n, y_2^n) respectively, we have

$$\begin{aligned}
&\kappa u(t^n, s_1^n, y_1^n) - b_n - \tfrac{1}{2}\sigma^2 (s_1^n)^2 X_{11}^n + L(s_1^n, y_1^n, p^n, q^n) \leq 0 \\
&\kappa v(t^n, s_2^n, y_2^n) - b_n - \tfrac{1}{2}\sigma^2 (s_2^n)^2 Y_{11}^n + L(s_2^n, y_2^n, p^n, q^n) \geq 0,
\end{aligned}$$

where

$$L(t, s, y, p, q) := -B_1(y, e^{-\kappa t}p)q + \lambda(y) B_1(y, e^{-\kappa t}p)p - sB_2(y, e^{-\kappa t}p)p - e^{\kappa t} s f(y) B_3(y, e^{-\kappa t}p).$$

As a consequence,

$$\begin{aligned}0<\kappa\delta<\kappa(u(t^n,s_1^n,y_1^n)-v(t^n,s_2^n,y_2^n))\le\\ \le -\tfrac{1}{2}\sigma^2(s_2^n)^2Y_{11}^n+\tfrac{1}{2}\sigma^2(s_1^n)^2X_{11}^n+\\ +L(t^n,s_2^n,y_2^n,p^n,q^n)-L(t^n,s_1^n,y_1^n,p^n,q^n).\end{aligned}\tag{4.59}$$

On the other hand, (4.58) implies

$$\tfrac{1}{2}\sigma^2(s_1^n)^2X_{11}^n-\tfrac{1}{2}\sigma^2(s_2^n)^2Y_{11}^n\le\tfrac{3}{2}\sigma^2n(s_1^n-s_2^n)^2$$

that converges to 0 as $n\to\infty$ due to (4.57). Let us now analyze the difference $L(t^n,s_2^n,y_2^n,p^n,q^n)$ $-L(t^n,s_1^n,y_1^n,p^n,q^n)$. We have the following estimates for the corresponding terms, where C (resp. C_R) is a Lipschitz constant (depending on R), that changes from line to line

$$\begin{aligned}|B_1(y_1^n,e^{-\kappa t^n}p^n)q^n-B_1(y_2^n,e^{-\kappa t}p^n)q^n|&\le C|y_1^n-y_2^n||q_n|,\\ |\lambda(y_1^n)B_1(y_1^n,e^{-\kappa t^n}p^n)p^n-\lambda(y_2^n)B_1(y_2^n,e^{-\kappa t^n}p^n)p^n|&\le C|y_1^n-y_2^n||p^n|,\\ |s_1^nB_2(y_1^n,e^{-\kappa t^n}p^n)p^n-s_2^nB_2(y_2^n,e^{-\kappa t^n}p^n)p^n|&\le C|(s_1^n-s_2^n)p^n|+C_R|(y_1^n-y_2^n)p^n|,\\ |e^{\kappa t^n}s_1^nf(y_1^n)B_3(y_1^n,e^{-\kappa t^n}p^n)-e^{\kappa t^n}s_2^nf(y_2^n)B_3(y_2^n,e^{-\kappa t^n}p^n)|&\le C_R(|s_1^n-s_2^n|+|y_1^n-y_2^n|).\end{aligned}$$

Since all of the upper bounds converge to 0 as $n\to\infty$, the right-hand side in (4.59) is bounded by something that converges to 0 as $n\to\infty$. Hence we obtain a contradiction for large n. □

Because of lack of precise reference, we provide a comparison result also in the case of delta constraints leading to the variational inequality ($\mathbf{PDE}^\delta$).

Theorem 4.8.6. *Suppose that the resilience function h is Lipschitz continuous and Assumption 4.4.3 is in force. Let u (resp. v) be bounded upper- (resp. lower-) semicontinuous viscosity subsolution (resp. supersolution) of the variational inequality* ($\mathbf{PDE}^\delta$) *with the terminal condition* ($\mathbf{BC}^\delta$)*. Then $u\le v$ on $[0,T]\times\mathbb{R}_+\times\mathbb{R}$.*

Proof. We argue by contradiction. For any $a>0$, set $\mathcal{O}_a:=[a,\infty)\times[-1/a,1/a]$. Suppose that

$$\sup_{(t,s,y)\in[0,T]\times\mathbb{R}_+\times\mathbb{R}}(u-v)>0.$$

Then there exists some $a>0$ such that $\sup_{(t,s,y)\in[0,T]\times\mathcal{O}_a}(u-v)>0$. For $\kappa>0$, consider $\tilde u:=e^{\kappa t}u$ and $\tilde v:=e^{\kappa t}v$. Then $\tilde u$ (resp. $\tilde v$) is a viscosity sub- (resp. super-)solution of

$$\min\{\kappa\varphi+\tilde{\mathcal{L}}[\varphi],\mathcal{H}_{\mathcal{K},t}\varphi\}=0$$

with the boundary condition $\min\{\varphi(T,\cdot)-H(\cdot),\mathcal{H}_{\mathcal{K},T}\varphi\}=0$, where

$$\tilde{\mathcal{L}}[\varphi](t,s,y)=-\partial_t\varphi+h(y+1/\lambda\log(\lambda e^{-t\kappa}\partial_s\varphi+1))\partial_y\varphi-1/2\sigma^2s^2\partial_{ss}\varphi$$

and $\mathcal{H}_{\mathcal{K},t}\varphi = \lambda e^{-\kappa t}\partial_s\varphi + 1 - e^{-\lambda K}$ for $t \in [0,T]$.

Consider

$$\Theta_n := \sup_{(t,x_1,x_2)\in[0,T]\times\mathcal{O}_a^2} \tilde{u}(t,x_1) - \tilde{v}(t,x_2) - \frac{n}{2}|x_1 - x_2|^2 - \frac{1}{2n}|x_1|^2.$$

We have $\Theta_n > \iota$ for some $\iota > 0$. Since $\tilde{u} - \tilde{v}$ is upper-semicontinuous, it attains its maximum on the compact set $[0,T]\times\mathcal{O}_a^2$ at $(t_n, x_1^n, x_2^n) \in [0,T]\times\mathcal{O}_a^2$. By the arguments in the proof of [BLZ16, Lemma 3.11], after possibly passing to a subsequence we have

1. $\lim_{n\to\infty}\Theta_n = \sup_{(t,s,y)\in[0,T]\times\mathcal{O}_a}(\tilde{v} - \tilde{u}) \geq \iota > 0$,

2. $n|x_1^n - x_2^n|^2 \to 0$ and $\frac{1}{n}|x_1|^2 \to 0$ as $n \to \infty$.

Note that also

$$\lim_{n\to\infty} \tilde{u}(t_n, x_1^n) - \tilde{v}(t_n, x_2^n) \geq \iota. \tag{4.60}$$

Case 1: Suppose, after passing to a subsequence, that $t_n = T$ for all n. Then Ishii's lemma together with the viscosity property of $\tilde{u}$ and $\tilde{v}$ give

$$\begin{aligned}&\min\left\{\tilde{u}(T,x_1^n) - H(x_1^n),\ \lambda e^{-\kappa T}(p_n + s_1^n/n) + 1 - e^{-\lambda K}\right\} \leq 0,\\ &\min\left\{\tilde{v}(T,x_2^n) - H(x_2^n),\ \lambda e^{-\kappa T}p_n + 1 - e^{-\lambda K}\right\} \geq 0,\end{aligned}$$

where $p_n = n(s_1^n - s_2^n)$. Hence we conclude that $\tilde{u}(T,x_1^n) \leq H(x_1^n)$ for all n. However, in this case since $\tilde{v}(T,x_2^n) \geq H(x_2^n)$ for all n we have

$$\tilde{v}(T,x_2^n) \geq H(x_2^n) \geq H(x_2^n) - H(x_1^n) + \tilde{u}(T,x_1^n),$$

which contradicts (4.60) for large n by continuity of H.

Case 2: We can now assume (after passing to a subsequence) that $t_n < T$ for all n. Set

$$p_n := n(s_1^n - s_2^n), \quad q_n := n(y_1^n - y_2^n).$$

By Ishii's lemma, see [CIL92, Theorem 8.3], using the viscosity property of $\tilde{u}$ and $\tilde{v}$, there exist $a_n \in \mathbb{R}$ and symmetric 2×2 matrices A_n, B_n (that satisfy a bound like in (4.58)) with

$$(a_n, (p_n + s_1^n/n, q_n), A_n) \in \bar{\mathcal{P}}^+_{\mathcal{O}_a}\bar{u}(t_n, x_1^n), \quad (a_n, (p_n, q_n), B_n) \in \bar{\mathcal{P}}^-_{\mathcal{O}_a}\bar{v}(t_n, x_2^n),$$

such that

$$\begin{aligned}\min\left\{-a_n + L(t_n, x_1^n, \tilde{u}(t_n,x_1^n), p_n + s_1^n/n, q_n, A_n),\ \lambda e^{-\kappa t_n}(p_n + s_1^n/n) + 1 - e^{-\lambda K}\right\} &\leq 0,\\ \min\left\{-a_n + L(t_n, x_2^n, \tilde{v}(t_n,x_2^n), p_n, q_n, B_n),\ \lambda e^{-\kappa t_n}p_n + 1 - e^{-\lambda K}\right\} &\geq 0,\end{aligned}$$

where for $t \in [0,T]$, $x = (x_1, y_1) \in \mathbb{R}^2$, $\ell, p, q \in \mathbb{R}$ and a 2×2 matrix A

$$L(t, x = (x_1, y_1), \ell, p, q, A) := \kappa\ell + h(y_1 + 1/\lambda \log(\lambda e^{-\kappa t} p + 1))q - 1/2\sigma^2 x_1^2 A_{11}.$$

Therefore, we have

$$-a_n + L(t_n, x_1^n, \tilde{u}(t_n, x_1^n), p_n + s_1^n/n, q_n, A_n) \leq 0.$$

Note also that on the set $\{(t, y, p) \in [0,T] \times \mathbb{R} \times \mathbb{R} \mid \lambda e^{-\kappa t} p + 1 - e^{-\lambda K} \geq 0\}$, the function

$$(t, y, p) \mapsto h(y + 1/\lambda \log(\lambda e^{-\kappa t} p + 1))$$

is Lipschitz continuous. Thus, we can argue like in the proof of Theorem 4.8.5 to derive a contradiction as follows: one gets the estimate

$$\kappa(\tilde{u}(t_n, x_1^n) - \tilde{v}(t_n, x_2^n)) \leq C \left(n|x_1^n - x_2^n|^2 + q_n s_1^n/n\right)$$

for some constant $C > 0$ that does not depend on n, hence the right-hand side converges to 0 as $n \to \infty$, contradicting $\lim_{n\to\infty} \Theta_n \geq \iota > 0$. □

Remark 4.8.7. By Theorem 4.8.2 and Theorem 4.8.3 we know that w_* (resp. w^*) is a supersolution (subsolution) of (**PDE**$^\delta$) with boundary condition (**BC**$^\delta$) and hence Theorem 4.8.6 gives that $w_* \geq w^*$ on $[0,T] \times \mathbb{R}_+ \times \mathbb{R}$. However, by definition it is clear that $w_* \leq w^*$ and hence we have the $w_* = w^*$ on $[0,T] \times \mathbb{R}_+ \times \mathbb{R}$. On the other hand, $w_* \leq w \leq w^*$ on $[0,T) \times \mathbb{R}_+ \times \mathbb{R}$. To obtain equality also for $t = T$, note that the super-/sub-solution property of w_*/w^* respectively implies also $w_*(T, \cdot) \geq H(\cdot)$ and $w_*(T, \cdot) \leq H(\cdot)$, hence the T-value of w_* is exactly H. Since also $H(\cdot) = w(T, \cdot)$ by definition, we conclude the equality $w_* = w^* = w$ also on $\{T\} \times \mathbb{R}_+ \times \mathbb{R}$.

The same conclusion holds for (**PDE**) with the boundary condition (**BC**).

5 Cross-impact and hedging in multi-asset price impact models

In this chapter, we consider general multi-asset price impact models with both transient and permanent impact that we define in Section 5.1. We derive structural conditions in Theorem 5.1.4 on the price specification that prevent the existence of trivial arbitrages in the sense of Definition 5.1.2. These structural conditions moreover allow us to identify the asymptotically realizable proceeds for a large set of trading strategies, see (5.10) and Theorem 5.2.1, that includes in particular all semimartingales. In an additive impact specification (Section 5.3.1), we consider the problem of pricing and hedging of non-covered options, defined in Section 5.3, and derive the pricing pde characterizing the minimal superhedging prices in Section 5.3.3. The technical proofs related to our application of hedging are delegated to Section 5.3.4.

General notations. For $n \in \mathbb{N}$, we identify $\mathbb{R}^n = \mathbb{R}^{n\times 1}$, i.e. vectors are column vectors, and for $x \in \mathbb{R}^n$ we identify its coordinates as $(x^1, \dots, x^n)$. On $\mathbb{R}^n$ we consider the norm $|x| := \sup_{i=1,\dots,n} |x^i|$ (and similarly on $\mathbb{R}^{n\times m}$), while on the space of càdlàg function of finite variation we denote the finite-variation norm (w.r.t. $|\cdot|$) by $|\cdot|_{\mathrm{TV}}$. For a function φ that depends on the argument $x \in \mathbb{R}^n$, we use the notation $\mathrm{grad}_x\varphi = \partial_x\varphi$ for $(\frac{\partial\varphi}{\partial x^i})_{i=1}^n$ and note that this is a row vector; similarly $\partial^2_{xy}\varphi$ will denote the matrix of all cross second order derivatives in its (x, y)-argument. For vectors $x, y \in \mathbb{R}^n$, $\langle x, y\rangle := \sum_{i=1}^n x^i y^i$ is the Eucledian inner product, while for $\mathbb{R}^n$-valued semimartingales X and predictable integrands ϑ $\int\langle\vartheta_u, \mathrm{d}X_u\rangle := \sum_{i=1}^d \int \vartheta^i_u \,\mathrm{d}X^i_u$, where the integrals are understood in Itô's sense. Moreover, for $\mathbb{R}^{1\times d}$-valued predictable integrands ϑ we also set $\int \vartheta_u \,\mathrm{d}X_u := \sum_{i=1}^n \int \vartheta^i_u \,\mathrm{d}X^i_u$.

5.1 Multi-asset models: the price impact function

Our mathematical framework is given by a measurable space $(\Omega, \mathcal{F})$ and a filtration $\mathbb{F} = (\mathcal{F}_t)_{t\geq 0}$. For our application of (multi-asset) price impact models, one typically specifies a fundamental price process $\overline{S}$ capturing the exogenous risks. At this point we consider general M-valued processes $\overline{S}$ and require that they have continuous paths, where $M \subseteq \mathbb{R}^d$. Latter we will need in addition that $\overline{S}$ is at least a $\mathbb{P}$-semimartingale for some probability measure $\mathbb{P}$ on $(\Omega, \mathcal{F})$, but for the results in this section a probabilistic structure will not be needed and only continuity of the paths of $\overline{S}$ will be used. Examples for M include $M = \mathbb{R}^d$ for very general setups, or $M = \mathbb{R}^d_+ = (0, \infty)^d$ in case $\overline{S}$ is the unaffected price process of assets that have positive prices, like for the (single-asset) multiplicative impact models considered so far.

To model the impact on prices from the trading actions of a large trader, let the $\mathbb{R}^d$-valued process Θ denote the evolution of her holdings in the risky assets, and consider the impact process $Y = Y^{\Theta}$ that evolves according to

$$\mathrm{d}Y_t = -h(Y_t)\,\mathrm{d}t + A\,\mathrm{d}\Theta_t, \quad Y_{0-} = y_0 \in \mathbb{R}^d, \tag{5.1}$$

where A is a $d \times d$ invertible matrix, and $h : \mathbb{R}^d \to \mathbb{R}^d$ is Lipschitz continuous. A typical example for h is $h(y) = By$, where B is a diagonal matrix with diagonal entries $\beta_1, \dots, \beta_d \geq 0$ corresponding to exponential resilience in each component of the impact process. The function h can be more general but it should model the resilience effects of impact. The exact form of h will be immaterial for our analysis in this chapter but we will make the following assumption:

$$\text{For each bounded } \Theta \text{ the process } Y^{\Theta} \text{ is bounded.} \tag{A1}$$

We consider models that combine both temporary and permanent impact. The process Y models the transient component of the price impact while the holdings in the risky asset gives the permanent impact. In this sense, the prices of risky assets are a function of some exogenously given risk factor process $\overline{S}$ and the processes Y and Θ, i.e. we postulate that for $i = 1, \dots, d$, and $t \geq 0$

$$\text{Price of asset } i \text{ is } S^i_t = g^i(\overline{S}_t, Y_t, \Theta_t),$$

where $g^i : M \times \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is a locally Lipschitz continuous function, and prices are understood in discounted units of a riskless asset ("cash") which serves as a numeraire. Thus, the (affected by the large trader's actions) price of the risky assets is $S = g(\overline{S}, Y, \Theta)$, where $g := (g^1, \dots g^d)$ is the *price impact function.* To summarize, impacted prices are determined (through the price impact function) by the process $\overline{S}$, the initial level of impact Y_{0-} and the trading strategy Θ (that also drives (5.1)).

Remark 5.1.1. Cross-impact in the volume imbalances modelled by Y via the matrix A could be motivated as follows, cf. [TWG17]. A buy order in one asset might lead to cancellation of orders in another asset that would change its price. Let us stress that in our model specification we have zero bid-ask spread (one risky asset price for both buying and selling) and thus the gaps in the LOBs resulting from execution or cancellations (due to cross-impact) are instantaneously filled in from the opposite side. However, the mean-reverting property of Y renders these shifts in the demand/supply imbalances transient, i.e. they will eventually recover in time to a neutral state (a zero of the resilience function h). Thus, the Y process captures the transient impact and cross-impact, while having the additional dependence on Θ in the prices S allows also for permanent cross-impact component.

Having specified the price process, next we would like to define proceeds from trading for general strategies (including jumps and even of infinite variation), that would later allow us to define also the wealth process. The starting point of our analysis is that similarly to the discussion in the beginning of Section 2.2 (cf. Lemma 2.2.1), the gains from trading for absolutely continuous strategy Θ should be

$$L(\Theta) = -\int_0^{\cdot} \langle S_u, \, \mathrm{d}\Theta_u \rangle = -\int_0^{\cdot} \langle g(\overline{S}_u, Y_u, \Theta), \, \mathrm{d}\Theta_u \rangle. \tag{5.2}$$

In the multi-asset setup, there are many ways how one could approximate a block trade by a sequence of absolutely continuous trades. For instance, a block trade in two assets can be approximated by quickly trading first in the first asset and afterwards in the second asset, or first trading in the second asset and afterwards in the first. A sensible model specification should give that such different approximations would not yield completely different proceeds because otherwise one could easily build quick round trips that yield as much proceeds as one wants. To make this precise, a model specification should not allow for *profitable asymptotically instantaneous round trips* in the sense of Definition 5.1.2, and in particular different ways of approximating block trades in short time should give the same proceeds/costs, at least in the limit when the time for realization converges to 0.

Definition 5.1.2. *A sequence of* (Θ^n) *of absolutely continuous round trips that are completed in time* $1/n$*, i.e.* $\Theta^n_t = \Theta^n_{0-}$ *for* $t \geq 1/n$*, and that is of bounded total variation, i.e.* $\sup_n |\Theta^n(\omega)|_{TV} < \infty$ *for all* $\omega \in \Omega$*, is called* asymptotically instantaneous round trip*. It is* always profitable *if also*

$$\limsup_{n\to\infty} L_{1/n}(\Theta^n)(\omega) > 0 \quad \forall \omega \in \Omega.$$

We say that the price impact function g does not allow for occasionally profitable asymptotically instantaneous round trips *if for every specification of* $\overline{S}$*,* Y_{0-} *and* Θ_{0-}*, and any asymptotically instantaneous round trip* (Θ^n)*, it holds*

$$\limsup_{n\to\infty} L_{1/n}(\Theta^n)(\omega) = 0 \quad \forall \omega \in \Omega.$$

Remark 5.1.3. Our notion of asymptotically instantaneous round trips is defined at initial time $t = 0$. However, as one can easily see from the proof of Theorem 5.1.4 (being pathwise), the structural condition derived there is sufficient to rule out occasionally profitable asymptotically instantaneous round trips that might start at any random time τ. More precisely, for any sequence $(\Theta^n)_{n\in\mathbb{N}}$ of absolutely continuous processes with paths of bounded variation for which there exists a finite random time τ so that for all $n \in \mathbb{N}$, $\Theta^n = \Theta^1$ on $[\![0, \tau]\!]$ and $\Theta^n_\tau = \Theta^n_{\tau+t}$ for $t \geq 1/n$, we have

$$\lim_{n\to\infty} L_{\tau+1/n}(\Theta^n)(\omega) - L_\tau(\Theta^n)(\omega) = 0 \qquad \forall \omega \in \Omega.$$

The following result gives a structural condition on the price function g that rules out profitable asymptotically instantaneous round trips.

Theorem 5.1.4. *The price impact function* g *does not allow for occasionally profitable asymptotically instantaneous round trips if and only if for every* $\bar{s} \in M$*,* $y, \theta \in \mathbb{R}^d$*, there* *exists a continuously differentiable function* $G^{\bar{s},y,\theta}: \mathbb{R}^d \to \mathbb{R}$ *such that*

$$\frac{\partial G^{\bar{s},y,\theta}}{\partial x}(x) = g(\bar{s}, y + Ax, \theta + x)^{tr} \quad \forall x \in \mathbb{R}^d.$$

Moreover, if this condition is violated for some $\bar{s} \in M$*,* $y, \theta \in \mathbb{R}^d$*, then we can find a deterministic sequence of asymptotically instantaneous round trips that are always profitable for every specification of* $\overline{S}$ *with* $\overline{S}_0 = \bar{s}$*.*

Proof. If the map $x \mapsto g(\bar{s}, y+Ax, \theta+x)$ is not a gradient field, we can find a piecewise-C^1 curve $\gamma : [0,1] \to \mathbb{R}^d$ with $\gamma(0) = \gamma(1) = 0$ such that

$$-\int_0^1 \langle g(\bar{s}, y + A\gamma(u), \theta + \gamma(u)), \, \mathrm{d}\gamma(u)\rangle > 0. \tag{5.3}$$

Indeed, in this case there exist $x \in \mathbb{R}^d$ and two piecewise-C^1 curves $\gamma_{1,2} : [0, 1/2] \to \mathbb{R}^d$ from 0 to x along which the integrated vector field yields different values, i.e.

$$\int_{\gamma_1} \langle g(\bar{s}, y + A\gamma_1(u), \theta + \gamma_1(u)), \, \mathrm{d}\gamma_1(u)\rangle < \int_{\gamma_2} \langle g(\bar{s}, y + A\gamma_2(u), \theta + \gamma_2(u)), \, \mathrm{d}\gamma_2(u)\rangle.$$

Thus, $\gamma(t) := \gamma_1(t)\mathbb{1}_{[0,1/2]}(t) + \gamma_2(1-t)\mathbb{1}_{(1/2,1]}(t)$ is a piecewise-C^1 closed loop starting at 0 such that (5.3) holds. Now for any market model with $\overline{S}_0 = \bar{s}$, $Y_0 = y$ and $\Theta_0 = \theta$, consider the round-trip strategies Θ^n defined by $\Theta^n_t := \gamma(nt)$ for $t \in [0, 1/n]$, and constant for $t \geq 1/n$. After integration by substitution, we get for the proceeds

$$\begin{aligned} L_{1/n}(\Theta^n) &= -\int_0^1 \langle g(\bar{s} + \varepsilon_1^n(u), y + \varepsilon_2^n(u) + A\gamma(u), \theta + \gamma(u)), \;\; \mathrm{d}\gamma(u)\rangle \\ &= -\int_0^1 \langle g(\bar{s}, y + A\gamma(u), \theta + \gamma(u)), \;\; \mathrm{d}\gamma(u)\rangle + \Xi_n, \end{aligned} \tag{5.4}$$

where $\varepsilon_1^n(x) = \overline{S}_{x/n} - \overline{S}_0$, $\varepsilon_2^n(x) = -\int_0^{x/n} h(Y_u^{\Theta^n})\,\mathrm{d}u$ and

$$\Xi_n := \int_0^1 \langle g(\bar{s}, y + A\gamma(u), \theta + \gamma(u)) - g(\bar{s} + \varepsilon_1^n(u), y + \varepsilon_2^n(u) + A\gamma(u), \theta + \gamma(u)), \;\; \mathrm{d}\gamma(u)\rangle.$$

In particular, $\Xi_n(\omega) \to 0$ as $n \to \infty$ for every $\omega \in \Omega$ by continuity of g and dominated convergence. Therefore, we have $\limsup_n L_{1/n}(\Theta^n)(\omega) > 0$ for every $\omega \in \Omega$, and thus the market impact function allows profitable asymptotically instantaneous round trips.

Now we argue the other direction and assume the existence of potential functions $G^{\bar{s},y,\theta}$. Let Θ^n be an arbitrary sequence of absolutely continuous round trips executing in time $1/n$ such that $\sup_n |\Theta^n(\omega)|_{\mathrm{TV}} < \infty$ for all $\omega \in \Omega$. We have (using the gradient field structure for g for the second equality)

$$L_{1/n}(\Theta^n) = -\int_0^1 \langle g(\bar{s}, y + A\gamma^n(u), \theta + \gamma^n(u)), \;\; \mathrm{d}\gamma^n(u)\rangle + \Xi_n = \Xi_n,$$

where $\gamma^n(t) := \Theta^n_{t/n}$ for $t \in [0,1]$, $n \in \mathbb{N}$,

$$\Xi_n := \int_0^1 \langle g(\bar{s}, y+A\gamma^n(u), \theta+\gamma^n(u)) - g(\bar{s}+\varepsilon_1^n(u), y+\varepsilon_2^n(u)+A\gamma^n(u), \theta+\gamma^n(u)),\ \mathrm{d}\gamma^n(u)\rangle,$$

and $\varepsilon_1^n(x) = \overline{S}_{x/n} - \overline{S}_0$, $\varepsilon_2^n(x) = -\int_0^{x/n} h(Y_u^{\Theta^n})\,\mathrm{d}u$. Here we again have $\Xi_n(\omega) \to 0$ for all $\omega \in \Omega$. Indeed, with

$$g_u^n := g(\bar{s}, y+A\gamma^n(u), \theta+\gamma^n(u)) - g(\bar{s}+\varepsilon_1^n(u), y+\varepsilon_2^n(u)+A\gamma^n(u), \theta+\gamma^n(u)) \text{ for } u \in [0,1],$$

we have

$$|\Xi_n| \le \left(\sup_{0\le t\le 1} |g_t^n|\right) |\gamma^n|_{TV} \le \left(\sup_{0\le t\le 1} |g_t^n|\right) \sup_k |\gamma^k|_{TV} \to 0 \quad \text{for all } \omega \in \Omega,$$

because by the local Lipschitz property of g we have for some constant $C = C(\omega)$

$$\sup_{0\le t\le 1} |g_t^n| \le C\Big(\sup_{0\le t\le 1} |\varepsilon_1^n(t)| + \sup_{0\le t\le 1} |\varepsilon_2^n(t)|\Big) \to 0 \text{ as } n \to \infty.$$

Therefore, for all $\omega \in \Omega$ we have $\Xi_n(\omega) \to 0$ as $n \to \infty$, and in particular

$$\lim_{n\to\infty} L_{1/n}(\Theta^n)(\omega) = 0 \quad \forall \omega \in \Omega.$$

□

Remark 5.1.5. Now we discuss the role of A in the gradient field condition from Theorem 5.1.4. In general the existence of a potential function G for the vector field $y \mapsto g(\bar{s}, Ay)$ does not imply the existence of a potential for the vector field $y \mapsto g(\bar{s}, y)$. Indeed, let us suppress the dependence in $\bar{s}$ for simplicity of notation, consider $d = 2$ and assume that g is continuously differentiable. For a matrix $A = (a\ b; c\ d)$, the vector field $(f_1, f_2)(y_1, y_2) := (g_1, g_2)(ay_1 + by_2, cy_1 + dy_2)$ is a gradient field if and only if $\partial_{y_2} f_1(y_1, y_2) = \partial_{y_1} f_2(y_1, y_2)$ for all $(y_1, y_2) \in \mathbb{R}^2$, which reads

$$(b\partial_{y_1} g_1 + d\partial_{y_2} g_1)(ay_1 + by_2, cy_1 + dy_2) = (a\partial_{y_1} g_2 + c\partial_{y_2} g_2)(ay_1 + by_2, cy_1 + dy_2).$$

Now for $b = -d$ it's easy to see that $(g_1, g_2)(y_1, y_2) = (y_1 + y_2, 1)$ satisfies the above equality but is not a gradient field, since $\partial_{y_2} g_1 \neq \partial_{y_1} g_2$.

Let us however point out that if $d = 1$ or $A = aI_d$ with $a \neq 0$, then existence of a potential for $y \mapsto g(\bar{s}, y)$ is equivalent to the existence of a potential for $y \mapsto g(\bar{s}, Ay)$.

From now on we assume that the multi-asset market impact model is free of occasionally profitable asymptotically instantaneous round trips, or equivalently as we just showed in Theorem 5.1.4

$$\text{for all } \bar{s}, y, \theta \in \mathbb{R}^d \text{ there exists a potential function } G^{\bar{s},y,\theta} \text{ from Theorem 5.1.4.} \quad \text{(A2)}$$

To determine how these potential functions $G^{\bar{s},y,\theta}$ are related for different $\bar{s}, y, \theta$, note that for each $(\bar{s}, y, \theta)$, $G^{\bar{s},y,\theta}$ is unique if we fix $G^{\bar{s},y,\theta}(0)$. Thus without loss of generality we assume that $G^{\bar{s},y,\theta}(0) = 0$ for all $\bar{s} \in M$, $y, \theta \in \mathbb{R}^d$. In this case moreover, the costs[1] from a block trade of size Δ, given as the limit of approximating continuous tradings in short time intervals, will be simply $G^{\bar{s},y,\theta}(\Delta)$, provided that the state of $(\overline{S}, Y, \Theta)$ before the jump is $(\bar{s}, y, \theta)$. Absence of profitable asymptotically instantaneous round trips implies also that splitting a block trade into two or more block trades executed one after the other should not make a difference in terms of proceeds obtained. This in particular gives a representation of $G^{\bar{s},y,\theta}$ in terms of $(G^{\bar{s},0,\theta})_{\theta\in\mathbb{R}^d}$. Indeed, splitting every block trade of size $x \in \mathbb{R}^d$ into a block trade of size $-A^{-1}y$ and then immediately a block trade of size $A^{-1}y + x$ gives

$$G^{\bar{s},y,\theta}(x) = G^{\bar{s},y,\theta}(-A^{-1}y) + G^{\bar{s},0,\theta-A^{-1}y}(A^{-1}y + x) \quad \forall x \in \mathbb{R}^d. \tag{5.5}$$

Thus, we can reduce the number of parameters needed to describe the family $\{G^{\bar{s},y,\theta}(x) \mid \bar{s}, y, \theta, x \in \mathbb{R}^d\}$. Indeed, consider the function $G : \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ defined by

$$G(\bar{s}, y, \theta) := G^{\bar{s},0,\theta-A^{-1}y}(A^{-1}y) \quad \forall \bar{s}, y, \theta \in \mathbb{R}^d. \tag{5.6}$$

Then

$$G(\bar{s}, y + Ax, \theta + x) = G^{\bar{s},0,\theta-A^{-1}y}(A^{-1}y + x) \quad \forall x \in \mathbb{R}^d. \tag{5.7}$$

The advantage of considering the function G is that it depends only on the parameters $\bar{s}, y$, and θ and no more on x. Moreover, the function G encodes completely the information on block trades from $G^{\bar{s},y,\theta}$ in the following sense: a block trade of size x executed at state $(\bar{s}, y, \theta)$, imposes the costs (negative proceeds)

$$G(\bar{s}, y + Ax, \theta + x) - G(\bar{s}, y, \theta). \tag{5.8}$$

This could be seen directly by using the definition of G in (5.6) and (5.5), since $G^{\bar{s},y,\theta}(-A^{-1}y) = -G^{\bar{s},0,\theta-A^{-1}y}(A^{-1}y)$. The function G will play a crucial role in the subsequent analysis to identify the proceeds from general strategies (possibly of infinite variation), see (5.10) and Theorem 5.2.1 below.

We now end this section with different examples of (multi-asset) price impact models.

Example 5.1.6. 1. *Additive cross-impact models:* for $i = 1, \dots d$, let

$$g^i(\overline{S}, Y, \Theta) = \overline{S}^i + \sum_{j=1}^{d} \lambda_{ij} Y^j + \sum_{j=1}^{d} \gamma_{ij} \Theta^j,$$

where $\lambda_{ij}, \gamma_{ij} \in \mathbb{R}$. This model extends the one-dimensional setup of Obizhaeva-Wang ([OW13]) and is similar to the model in [GP16]. In this case, cross-impact is due to the matrices $\Lambda = (\lambda_{ij})_{0\le i,j\le d}$ and A for the transient impact and

[1] Negative costs should be understood as proceeds/gains from the trade

$\Gamma = (\gamma_{ij})_{0\le i,j\le d}$ for the permanent impact. In this case, Theorem 5.1.4 implies that the additive impact model is free of profitable asymptotically instantaneous round trips if and only if $x \mapsto (\Lambda A + \Gamma)x$ is a gradient field. Note that this is the case if and only if the matrix $(\Lambda A + \Gamma)$ is symmetric. Let us also point out that this is in line with the results in [HS04] that show in a discrete-time setup with only permanent impact that absence of price manipulations, being dynamic round trips with negative expected costs, requires symmetric cross-impact function. In our case with $h = 0$, we have precisely that $\Lambda A + \Gamma$ is the permanent cross-impact component.

Note that $G^{\bar{s},y,\theta}(x) = (\bar{s} + \Lambda y + \Gamma\theta)^{\mathrm{tr}} x + 1/2 x^{\mathrm{tr}}(\Lambda A + \Gamma)x$ and thus a block trade of size Δ imposes the costs $G^{\bar{s},y,\theta}(\Delta)$. Moreover,

$$G(\bar{s}, y, \theta) = (\bar{s} + \Gamma\theta - \Gamma A^{-1}y)^{\mathrm{tr}} A^{-1}y + 1/2(A^{-1}y)^{\mathrm{tr}}(\Lambda A + \Gamma)A^{-1}y. \tag{5.9}$$

2. *Multiplicative cross-impact model:* Let $A = I_d$ and $\lambda_i \in \mathbb{R}^d$ for $i = 1, \dots d$. Consider the market model with pure transient impact where the price processes are derived from the potentials

$$G^{\bar{s},y}(x) = \sum_{i=1}^{d} \bar{s}^i \left(\exp\left(\langle \lambda_i, y + x\rangle\right) - \exp\left(\langle \lambda_i, y\rangle\right)\right).$$

In this case the price processes will be

$$g^i(\bar{s}, y) = \sum_{j=1}^{d} \lambda_{ji} \bar{s}^j \exp\left(\langle \lambda_j, y\rangle\right).$$

The cross-impact is modeled by the matrix $\Lambda = (\lambda_{ij})_{1\le i,j\le d}$. Small entries λ_{ji} means small impact of asset j on the price of asset i. Note however that Λ need not be symmetric in this case. For diagonal matrix Λ we get a multi-asset model with exponential impact functions, generalizing the single-asset case analyzed in Section 4.4.2.

3. *Single-asset models.* In the setup of one risky asset ($d = 1$), one can always construct the potentials $G^{\bar{s},y,\theta}$, in contrast to the case $d > 1$. Indeed, in the one-dimensional setup we can simply integrate g to get for $\bar{s}, y, \theta \in \mathbb{R}$

$$G^{\bar{s},y,\theta}(x) = \int_0^x g(\bar{s}, y + Au, \theta + u)\,\mathrm{d}u.$$

Note that the function G from equation (5.6) coincides with the function G from Section 2.4.5 (where we considered $A = 1$) that was crucial for deriving the asymptotically realizable proceeds there. This is actually not a coincidence. As we will also see in the general multi-asset case, the function G will enable us to identify the asymptotically realizable proceeds for general trading strategies.

5.2 The proceeds functional

In this section we generalize the definition of the proceeds functional L by continuous extension from simple trading strategies to general càdlàg strategies. Since our analysis relies on Itô calculus, we introduce at this stage a probability measure in our setup and assume additional structure on the fundamental price process $\overline{S}$. More precisely, let $\mathbb{P}$ be a measure on $(\Omega, \mathcal{F})$ and suppose that the filtration $\mathbb{F} = (\mathcal{F}_t)_{t\geq 0}$ satisfies the usual conditions of right-continuity and completeness. In particular, all semimartingales with respect to the filtration $\mathbb{F}$ are taken to have càdlàg paths. The fundamental price process $\overline{S}$ is moreover assumed to be a continuous $(\mathbb{P}, \mathbb{F})$-semimartingale. In addition, for every $i, j \in \{1, \dots, d\}$ the quadratic (co-)variation process $\langle \overline{S}^i, \overline{S}^j \rangle$ is absolutely continuous with respect to the Lebesgue measure, meaning that for all $i, j \in \{1, \dots, d\}$ there exist predictable processes α^{ij} such that $\mathrm{d}\langle \overline{S}^i, \overline{S}^j \rangle_t = \alpha^{ij}_t \,\mathrm{d}t$. Moreover, α^{ij} are assumed to be $\mathbb{P}$-a.s. bounded on any compact time interval $[0, T]$.

Having the gradient field structure for the price impact function not only prevents profitable asymptotically instantaneous round trips, but also allows us to rewrite the proceeds functional (5.2) as follows. In what follows, we also assume that

$$G \in C^{2,1,1}, \tag{A3}$$

i.e. it is twice continuously differentiable in the $\bar{s}$ argument and once continuously differentiable in the y and θ arguments. Then differentiating in x in (5.7) and setting $x = 0$ gives in particular $\mathrm{grad}_y G(\bar{s}, y, \theta)\ A + \mathrm{grad}_\theta G(\bar{s}, y, \theta) = g(\bar{s}, y, \theta)^{\mathrm{tr}}$. Thus, after integration by parts one can rewrite the proceeds for all continuous finite-variation Θ from (5.2) in the following equivalent way:

$$\begin{aligned} L_t(\Theta) = &\int_0^t \mathrm{grad}_{\bar{s}} G(\overline{S}_u, Y_u, \Theta_u)\,\mathrm{d}\overline{S}_u - \int_0^t \mathrm{grad}_y G(\overline{S}_u, Y_u, \Theta_u) h(Y_u)\,\mathrm{d}u \\ &+ \frac{1}{2} \sum_{i,j=1}^d \int_0^t \frac{\partial^2 G}{\partial \bar{s}^i \partial \bar{s}^j}(\overline{S}_u, Y_u, \Theta_u)\,\mathrm{d}\langle \overline{S}^i, \overline{S}^j \rangle_u - G(\overline{S}, Y, \Theta)\Big|_{0-}^{t}. \end{aligned} \tag{5.10}$$

Note that the right-hand side of (5.10) is well-defined for a larger class of controls Θ than absolutely continuous controls, in particular for all càdlàg adapted Θ. In fact, we have the following stability result that allows us to extend the definition of proceeds L to general càdlàg Θ. Let $\mathbb{D}_{\mathrm{ucp}}$ (resp. $\mathbb{D}_{\mathrm{ucp,b}}$) be the space of adapted càdlàg processes (that are resp. bounded) endowed with the topology of uniform convergence on compacts in probability.

Theorem 5.2.1. *The functional $L : \mathbb{D}_{ucp,b} \to \mathbb{D}_{ucp}$ defined in* (5.10) *is continuous.*

Proof. Note that the map $\mathbb{D}_{\mathrm{ucp,b}} \to \mathbb{D}_{\mathrm{ucp,b}}$, $\Theta \mapsto Y^\Theta$, given by (5.1) is continuous; indeed, the straightforward multi-dimensional generalization of [PTW07, Thm. 4.1] gives ω-wise continuity (in the uniform topology on the space of càdlàg paths) which implies the claimed continuity (in the ucp topology), recall also equation (A1) for the reason why the solution map Y^Θ is bounded. Hence, the local Lipschitz continuity of $\mathrm{grad}_{\bar{s}}G$ implies that the map $\mathbb{D}_{\mathrm{ucp,b}} \to \mathbb{D}_{\mathrm{ucp}}$, $\Theta \mapsto \mathrm{grad}_{\bar{s}}G(\overline{S}, Y^{\Theta}, \Theta)$, is continuous. Therefore the stochastic integral term in (5.10) is continuous because of continuity of the stochastic integral map. The continuity of the two Riemann-Stieltjes terms in (5.10) follows from dominated convergence since ω-wise, on compact time intervals, the range of the integrands is bounded. Finally, continuity of the last term in (5.10) follows by the local Lipschitz continuity of G. □

Remark 5.2.2 (On extension to other topologies)**.** Let us consider instead of the uniform topology on the space of càdlàg paths (denoted by D) any other topology ρ for which the following holds: convergence of the paths of Θ^n to the paths of Θ in (D, ρ) implies that $Y_t^{\Theta^n} \to Y_t^{\Theta}$ at all continuity points t of Θ. The Skorokhod J_1 topology is such and we moreover suspect that one could extend the one-dimensional results in [PW10] to a multi-dimensional setup and show that the strong Skorokhod M_1 topology also satisfies this. Then adapting the arguments from the proof of Theorem 2.2.7 one can easily obtain the following stability result: $\Theta^n \to \Theta$ in (D, ρ), in probability, implies $L_t(\Theta^n) \to L_t(\Theta)$ at all continuity points t of Θ, in probability.

Having determined the proceeds for bounded càdlàg strategies, we can now define a notion of wealth for a self-financing portfolio that will be needed in the subsequent analysis.

Definition 5.2.3. *For a trading strategy $\Theta \in \mathbb{D}_{ucp,b}$ (in the risky assets),* the self-financing portfolio *(in the riskless and risky assets) with initial holdings in the riskless asset β_{0-} is the pair of processes (β, Θ) with*

$$\beta := \beta_{0-} + L(\Theta).$$

Here β denotes (the evolution of) the position in the riskless asset.

For a self-financing portfolio (β, Θ), the *instantaneous liquidation wealth* $V^{\mathrm{liq}} = V^{\mathrm{liq}}(\Theta)$ evaluates the holdings in risky assets by what the large trader would receive by an instantaneous block liquidation order, i.e. it is the process

$$V_t^{\mathrm{liq}}(\Theta) := \beta_t + G(\overline{S}_t, Y_t, \Theta_t) - G(\overline{S}_t, Y_t - A\Theta_t, 0) \quad \forall t \geq 0. \tag{5.11}$$

We end this section with the dynamics of V^{liq} in this multi-asset cross-impact model:

$$\begin{aligned} \mathrm{d}V_t^{\mathrm{liq}}(\Theta) =&\ \mathrm{d}(L_t(\Theta) + G(\overline{S}_t, Y_t, \Theta_t) - G(\overline{S}_t, Y_t - A\Theta_t, 0)) \\ &= (\mathrm{grad}_{\bar{s}}G(\overline{S}_t, Y_t, \Theta_t) - \mathrm{grad}_{\bar{s}}G(\overline{S}_t, Y_t - A\Theta_t, 0))\, \mathrm{d}\overline{S}_t \\ &\quad - (\mathrm{grad}_y G(\overline{S}_t, Y_t, \Theta_t) - \mathrm{grad}_y G(\overline{S}_t, Y_t - A\Theta_t, 0))h(Y_t)\, \mathrm{d}t \\ &\quad + \frac{1}{2} \sum_{i,j=1}^{d} \frac{\partial^2 G}{\partial \bar{s}^i \partial \bar{s}^j}(\overline{S}_t, Y_t, \Theta_t)\, \mathrm{d}\langle \overline{S}^i, \overline{S}^j \rangle_t. \end{aligned}$$

In what follows, we restrict our attention to additive impact for which this simplifies greatly.

5.3 Pricing and hedging with cross-impact

In this section, we consider the problem of pricing and hedging European contingent claims by the super-replication approach. We consider a specific multi-asset price impact model with additive structure that we specify in Section 5.3.1. In Section 5.3.2 we explain the notions of non-covered options, their hedging strategies and define the (minimal) superhedging price. To derive a characterization of the superhedging price, we formulate the pricing problem as a stochastic target problem, being a stochastic control problem with an almost sure constraint at terminal time (reflecting that hedging strategies super-replicate a payoff at maturity). For this stochastic target problem, we derive in Section 5.3.3 a dynamic programming principle after a suitable change of coordinates. Thereby, we show that the superhedging price is the viscosity solution of a semi-linear pde. We also construct hedging strategies when the pde admits sufficiently regular classical solution. The technical proofs are delegated for Section 5.3.4.

5.3.1 Additive cross-impact model

For the rest of this chapter we focus on the case of additive impact with both permanent and transient component from Example 5.1.6, 1. In this case, the price of the risky assets, when the evolution of the large trader's holdings are Θ, is

$$S = \overline{S} + \Lambda Y^{\Theta} + \Gamma\Theta,$$

where $\Lambda A + \Gamma$ is symmetric to prevent profitable asymptotically instantaneous round trips. Here the process $\overline{S}$ could be naturally interpreted as a fundamental (unaffected) price process that would prevail in absence of the large trader, i.e. in a liquid market. For its state space we take $M = \mathbb{R}^d$. Let $B = (B^1, \ldots, B^d)$ be a d-dimensional Brownian motion and assume that for $i = 1, \ldots, d$, and consider for the fundamental price process $\overline{S}$ the following specification:

$$\overline{S}^i_t = \overline{S}^i_0 + \int_0^t \mu^i_u \,\mathrm{d}u + \sum_{j=1}^d \sigma^{ij} B^j_t, \quad t \geq 0,$$

where μ^i is a bounded adapted process and $\sigma^{ij} \in \mathbb{R}$ for $j = 1, \ldots, d$. We also assume that the matrix $\Sigma = (\sigma^{ij})_{1 \leq i,j \leq d}$ is invertible. This condition is natural because it guarantees that there are no arbitrage opportunities for a frictionless model where prices are given by the fundamental price process, in particular for a "small investor" whose actions will not affect the price. To derive the dynamic programming principle in Section 5.3.3, we moreover assume that the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is complete with $\mathcal{F}$ being countably generated, cf. Section 4.8.1; for instance, it might be generated by the Brownian motion B and a sequence of random measures as in [ST02, Section 2.5].

Using the form of the function G from (5.9), we get directly the following surprisingly simple expressions:

$$\operatorname{grad}_{\bar{s}} G(\overline{S}_t, Y_t, \Theta_t) - \operatorname{grad}_{\bar{s}} G(\overline{S}_t, Y_t - A\Theta_t, 0) = \Theta_t^{\mathrm{tr}},$$
$$\operatorname{grad}_y G(\overline{S}_t, Y_t, \Theta_t) - \operatorname{grad}_y G(\overline{S}_t, Y_t - A\Theta_t, 0) = \Theta_t^{\mathrm{tr}}\Lambda,$$

and thus the dynamics of the instantaneous liquidation wealth takes the simple form

$$\mathrm{d}V_t^{\mathrm{liq}}(\Theta) = \Theta_t^{\mathrm{tr}}\,\mathrm{d}\overline{S}_t - \Theta_t^{\mathrm{tr}}\Lambda h(Y_t)\,\mathrm{d}t. \tag{5.12}$$

In particular, it is not difficult to show that this cross-impact model is free of arbitrages among the following set of admissible strategies

$$\mathcal{A}_{\mathrm{NA}} := \{(\Theta_t)_{t\geq 0} \mid \Theta \in \mathbb{D}_{\mathrm{ucp,b}} \text{ with } \Theta_{0-} = 0 \text{ and } \Theta_t = 0 \text{ on } [T,\infty) \text{ for some } T < \infty\}.$$

Proposition 5.3.1. *The market in free of arbitrage up to any finite time horizon $T \in [0,\infty)$ in the sense that there exists no $\Theta \in \mathcal{A}_{NA}$ with $\Theta_t = 0$ for $t \geq T$ such that for the self-financing strategy (β, Θ) with $\beta_{0-} = 0$ we have $\mathbb{P}[V_T^{liq} \geq 0] = 1$ and $\mathbb{P}[V_T^{liq} > 0] > 0$.*

Proof. For any such $\Theta \in \mathcal{A}_{\mathrm{NA}}$ that executes up to time T, the drift in (5.12) is absolutely continuous with bounded density, see (A1). Since Σ is invertible and Θ is bounded (and hence also Y by (A1)), Girsanov's theorem gives existence of a measure $\mathbb{P}^\Theta \approx \mathbb{P}$ such that V^{liq} is a $\mathbb{P}^\Theta$-martingale on $[0,T]$. In particular, $\mathbb{E}^{\mathbb{P}^\Theta}[V_T^{\mathrm{liq}}] = \mathbb{E}^{\mathbb{P}^\Theta}[V_0^{\mathrm{liq}}] = 0$, hence excluding the possibility of Θ being an arbitrage opportunity because $\mathbb{P}^\Theta$ and $\mathbb{P}$ are equivalent. ☐

5.3.2 European contingent claims and their superhedging prices

In this section, we adapt the notions from Sections 4.2 and 4.3.1 to the multi-asset case. In particular, we consider European contingent claims with fixed maturity $T \geq 0$ as in

Definition 5.3.2. *European contingent claim with maturity T is specified by a measurable map*

$$(s, y, \theta) \in \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d \mapsto (g_0(s, y, \theta), g_1(s, y, \theta)) \in \mathbb{R} \times \mathbb{R}^d$$

representing the payoff, where g_0 is the cash-settlement part and g_1 is the physical-delivery part at maturity. It entitles its holder the payment of $g_0(S_T, Y_T, \Theta_T)$ in cash and $g_1(S_T, Y_T, \Theta_T)$ in risky asset, where S_T, Y_T and Θ_T are respectively the risky asset price, the level of market impact and the risky asset position at maturity.

For the rest of this chapter we will fix the maturity $T \geq 0$. The seller of an (non-covered) European option with payoff (g_0, g_1) will have to hedge against possible losses due to her obligation to deliver the payoff at maturity. In order to obtain a dynamic programming principle later, we need to allow for jumps in the admissible hedging strategies. For this purpose, for $k \in \mathbb{N}$ let $\mathcal{U}_k$ denote the set of random $\{0,\ldots,k\}$-valued measures ν supported on $[-k,k]^d \times [0,T]$ that are adapted in the following sense: for every $A \in \mathcal{B}(\mathbb{R}^d)$, the process $t \mapsto \nu(A, [0,t])$ is adapted to the underlying filtration. Note that the elements of $\mathcal{U}_k$ have the representation

$$\nu(A, [0,t]) = \sum_{i=0}^{k} \mathbb{1}_{\{(\delta_i, \tau_i) \in A \times [0,t]\}},$$

where $0 \leq \tau_1 < \cdots < \tau_k \leq T$ are stopping times and δ_i is a $\mathbb{R}^d$-valued $\mathcal{F}_{\tau_i}$-random variable (might take values 0 as well). Consider also the set $\mathcal{U} := \bigcup_{k\geq 1} \mathcal{U}_k$ of all pure jump processes with bounded number of jumps that are of bounded size. The admissible trading strategies Θ that we will consider are bounded and are of the form

$$\Theta_t = \Theta_{0-} + \int_0^t a_s \,\mathrm{d}s + \int_0^t b_s \,\mathrm{d}B_s + \int_0^t \int_{\mathbb{R}^d} \delta \nu(\,\mathrm{d}\delta,\, \mathrm{d}s), \tag{5.13}$$

in which $\Theta_{0-} \in \mathbb{R}^d$, $\nu \in \mathcal{U}$ and $(a,b) \in \mathcal{A} := \bigcup_{k\geq 1} \mathcal{A}_k$, where for $k \in \mathbb{N}$

$$\mathcal{A}_k := \{(a,b) \mid a,b \text{ are } \mathbb{R}^d,\mathbb{R}^{d\times d}\text{-valued predict. with } |a| \vee |b| \leq k,\ \mathrm{d}t \otimes \mathrm{d}\mathbb{P}\text{-a.e.}\}.$$

We can identify the trading strategies by triplets $(a,b,\nu) \in \mathcal{A} \times \mathcal{U}$. For $k \in \mathbb{N}$, set

$$\Gamma_k := \{(a,b,\nu) \in \mathcal{A}_k \times \mathcal{U}_k \mid \Theta \text{ from (5.13) takes values in } [-k,k]^d\}$$

and let $\Gamma := \bigcup_{k\geq 1} \Gamma_k$.

Among the admissible trading strategies Γ, the following will be superhedging for a contingent claim with payoff (g_0, g_1).

Definition 5.3.3 (Hedging of non-covered option)**.** *A superhedging strategy is a self-financing strategy* (β, Θ) *with* $\Theta \in \Gamma$, $\Theta_{0-} = 0$, *and*

$$\beta_T \geq g_0(S_T, Y_T, \Theta_T) \quad \textit{and} \quad \Theta_T = g_1(S_T, Y_T, \Theta_T).$$

At this point we would like to stress the crucial assumption that a hedging strategy has to deliver at maturity exactly the physical part $g_1(S_T, Y_T, \Theta_T)$, and that any further (long or short) position in the underlying has to be unwound before the options are settled at the resulting price S_T, impact level Y_T and assets' position Θ. The minimal superhedging price is the infimum over all $p \in \mathbb{R}$ so that a superhedging strategy (β, Θ) with $\beta_{0-} = p$ exists.

To get a dynamic version of the superhedging problem, consider for

$$(t,z) = (t,s,y,\theta,v) \in [0,T] \times \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}$$

and $\gamma \in \Gamma$ the (dynamic version of) *the state process*

$$(Z_u^{t,z,\gamma})_{u\in[t,T]} = (S_u^{t,z,\gamma}, Y_u^{t,z,\gamma}, \Theta_u^{t,z,\gamma}, V_u^{\mathrm{liq},t,z,\gamma})_{u\in[t,T]},$$

where the processes $S^{t,z,\gamma}, Y^{t,z,\gamma}, \Theta^{t,z,\gamma}$ and $V^{\mathrm{liq},t,z,\gamma}$ correspond to the price, impact, risky asset position and instantaneous liquidation wealth processes on $[t,T]$ for the control $\Theta^{t,z,\gamma}$ associated with γ (from the decomposition (5.13) on $[t,T]$ instead), with initial condition (at time $t-$) being s, y, θ and v respectively. A strategy $\gamma \in \Gamma$ is superhedging if the state process at time T is (a.s.) in the set

$$\mathfrak{G} := \left\{(s,y,\theta,v) \in \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R} \mid \theta = g_1(s,y,\theta),\ v + G^{s-\Lambda y-\Gamma\theta,y,\theta}(-\theta) \geq g_0(s,y,\theta)\right\},$$

that we call the *target set.* Indeed, if $Z_T^{t,z,\gamma} \in \mathfrak{G}$, then $\Theta_T = g_1(S_T, Y_T, \Theta_T)$, i.e. the physical part will be delivered exactly, while since $V_T^{\mathrm{liq},t,z,\gamma} + G^{S_T-\Lambda Y_T-\Gamma\Theta_T,Y_T,\Theta_T}(-\Theta_T) = \beta_T$, there will be enough cash to meet the obligation stemming from the cash delivery part. All superhedging strategies for initial position θ in the risky assets are

$$\mathcal{G}(t,s,y,\theta,v) := \bigcup_{k\geq 1} \mathcal{G}_k(t,s,y,\theta,v)$$

with

$$\mathcal{G}_k(t,s,y,\theta,v) := \{\gamma \in \Gamma_k \mid Z_T^{t,s,y,\theta,v,\gamma} \in \mathfrak{G}\}.$$

Now we can define the minimal superhedging price as

$$w(t,s,y) := \inf_{k\geq 1} w_k(t,s,y), \quad \text{where} \quad w_k(t,s,y) := \inf\{v \mid \mathcal{G}_k(t,s,y,0,v) \neq \emptyset\}. \tag{5.14}$$

Note that the set of admissible superhedging strategies (identified with $\mathcal{G}(t,s,y,0,v)$) is a subset of $\mathcal{A}^{\mathrm{NA}}$, giving in particular that the minimal superhedging price of any payoff $(g_0, 0)$ with $g_0 > 0$ is strictly positive.

5.3.3 Characterization of the minimal superhedging price

In this section, we adapt the analysis from Section 4.3.2 to the current multi-asset setup. To derive a partial differential equation characterizing the minimal superhedging price w, we will rely on the following change of coordinates:

$$\mathcal{Y}(Y,\Theta) := Y - A\Theta, \qquad \mathcal{S}(S,Y,\Theta) := S - (\Lambda A + \Gamma)\Theta \quad (= \overline{S} + \Lambda(Y - A\Theta)). \tag{5.15}$$

The interpretation is that $\mathcal{Y}(Y,\Theta)$ is the impact that would prevail if we were to liquidate our position in the risky assets immediately, while $\mathcal{S}(S,Y,\Theta)$ is the price after instantaneous block trade of order $-\Theta$, given the price S and impact Y before that block trade. The reason why these processes will be relevant for the analysis later is the following dynamic programming principle, the proof of which is a straightforward adaptation of the proof of Theorem 4.3.1.

Theorem 5.3.4 (Geometric DPP)**.** *Fix* $(t,s,y,v) \in [0,T] \times \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}$.

1. *If $v > w(t,s,y)$, then there exists $\gamma \in \Gamma$ and $\theta \in \mathbb{R}^d$ such that*

$$V_\tau^{liq,t,z,\gamma} \geq w(\tau, \mathcal{S}(S_\tau^{t,z,\gamma}, Y_\tau^{t,z,\gamma}, \Theta_\tau^{t,z,\gamma}), Y_\tau^{t,z,\gamma} - A\Theta_\tau^{t,z,\gamma})$$

for all stopping times $\tau \geq t$, where $z = (\mathcal{S}(s,y,-\theta), y + A\theta, \theta, v)$.

2. *Let $k \geq 1$. If $v < w_{2k+2}(t,s,y)$, then for every $\gamma \in \Gamma_k$, $\theta \in [-k,k]^d$ and stopping time $\tau \geq t$ we have*

$$\mathbb{P}\left[\ V_\tau^{liq,t,z,\gamma} > w_k(\tau, \mathcal{S}(S_\tau^{t,z,\gamma}, Y_\tau^{t,z,\gamma}, \Theta_\tau^{t,z,\gamma}), Y_\tau^{t,z,\gamma} - A\Theta_\tau^{t,z,\gamma})\ \right] < 1$$

where $z = (\mathcal{S}(s,y,-\theta), y + A\theta, \theta, v)$.

The dynamic programming principle allows us to derive a pde for w. Indeed, for every smooth function $\varphi : [0,T] \times \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ we have

$$\begin{aligned} \mathrm{d}(V_t^{\mathrm{liq}}(\Theta) - \varphi(t, \mathcal{S}_t, \mathcal{Y}_t)) = {}& (\Theta_t^{\mathrm{tr}} - \partial_S \varphi(t, \mathcal{S}_t, \mathcal{Y}_t))\,\mathrm{d}\overline{S}_t \\ & - \{\varphi_t(t, \mathcal{S}_t, \mathcal{Y}_t) + \Theta_t^{\mathrm{tr}}\Lambda - (\partial_S\varphi\Lambda + \partial_Y\varphi)(t, \mathcal{S}_t, \mathcal{Y}_t)h(Y_t)\}\,\mathrm{d}t \\ & - \tfrac{1}{2}\sum_{i,j=1}^d (\Sigma\Sigma^{\mathrm{tr}})_{ij}\partial_{s^i s^j}\varphi(t, \mathcal{S}_t, \mathcal{Y}_t)\,\mathrm{d}t. \end{aligned} \tag{5.16}$$

So if the value function w is smooth enough, then the first part of Theorem 5.3.4 would give that an optimal strategy Θ must satisfy

$$\Theta_t = \partial_S w(t, \mathcal{S}_t, \mathcal{Y}_t)^{\mathrm{tr}}, \tag{5.17}$$

and that the drift in (5.16) should be non-negative. The second part of Theorem 5.3.4 on the other hand would imply that the drift cannot be strictly positive and thus we derive the following pde for w:

$$\mathcal{L}^{[\partial_S w]}w = \underbrace{-w_t - \tfrac{1}{2}\sum_{i,j=1}^d (\Sigma\Sigma^{\mathrm{tr}})_{ij}\partial^2_{s^i s^j} w}_{(I)} + \underbrace{\partial_Y w\ h(y + A(\partial_S w)^{\mathrm{tr}})}_{(II)} = 0, \tag{5.18}$$

where all derivatives of w above are evaluated at (t,s,y), and where for $\theta \in \mathbb{R}^{1\times d}$ and $\varphi \in C^{1,2,1}([0,\infty) \times \mathbb{R}^d \times \mathbb{R}^d)$ we have

$$\mathcal{L}^{[\theta]}\varphi(t,s,y) := -\varphi_t - \tfrac{1}{2}\sum_{i,j=1}^d (\Sigma\Sigma^{\mathrm{tr}})_{ij}\partial^2_{s^i s^j}\varphi - (\theta\Lambda - \partial_S\varphi\Lambda - \partial_Y\varphi)h(y + A\theta^{\mathrm{tr}}),$$

here again all the derivatives being evaluated at (t,s,y).

Note that the term (I) is linked to the dynamics of the fundamental price process and captures the correlated exogenous risks, while the non-linear term (II) is linked to the transient cross-impact nature. Thus, in full generality the cross-correlation and cross-impact are non-trivially coupled through the pde (5.18); see also Remark 5.3.8 for a discussion on how the pde simplifies when the terminal payoff is a function of S only. Note also, that the permanent cross-impact component is irrelevant for the pricing pde, but it will be relevant for the hedging strategy, see Remarks 5.3.7 and 5.3.8.

We now need to specify the boundary condition for the pricing pde. It is easy to get $w(T,s,y)$ since at time T the only possible superhedging strategy is to do a block trade (to deliver the physical part) and to cover the cash part of the payoff, in particular it is

$$H(s,y) := \inf\left\{g_0(s+\widetilde{\Gamma}\theta, y+A\theta,\theta) + G^{s-\Lambda y,y,0}(\theta) \mid \theta\in\mathbb{R}^d, \theta = g_1(s+\widetilde{\Gamma}\theta, y+A\theta,\theta)\right\},$$

where $\widetilde{\Gamma} := \Lambda A + \Gamma$ and we use the convention that $\inf\emptyset = +\infty$. For the analysis that follows we also need the functions H_n for $n\in\mathbb{N}$, where

$$H_n(s,y) := \inf\left\{g_0(s+\widetilde{\Gamma}\theta, y+A\theta,\theta) + G^{s-\Lambda y,y,0}(\theta) \mid \theta\in[-n,n]^d, \theta = g_1(s+\widetilde{\Gamma}\theta, y+A\theta,\theta)\right\}.$$

Note that $H = \inf_n H_n$. To summarize, in the view of the discussion so far, we expect w to be a solution of

$$\mathcal{L}^{[\partial_S\varphi]}\varphi \mathbb{1}_{[0,T[} + (\varphi - H)\mathbb{1}_{\{T\}} = 0 \quad \text{on } [0,T]\times\mathbb{R}^d\times\mathbb{R}^d. \tag{5.19}$$

In order to characterize the minimal superhedging price w as a viscosity solutions of the pricing pde (5.19), we need to work with the notion of discontinuous viscosity solution since *a priori* the function w is not continuous and in fact it is difficult to prove regularity of w directly. For this purpose, consider the relaxed semi-limits

$$w_*(t,s,y) := \liminf_{(t',s',y',k)\to(t,s,y,\infty)} w_k(t',s',y'), \tag{5.20}$$

$$w^*(t,s,y) := \limsup_{(t',s',y',k)\to(t,s,y,\infty)} w_k(t',s',y'), \tag{5.21}$$

where the limits are taken over $t' < T$.

For our main result in this section, we need the following assumption.

Assumption 5.3.5.

Bounded value function: *w_* and w^* are bounded on $[0,T]\times\mathbb{R}^d\times\mathbb{R}^d$;*

Regular payoff: *H is continuous, bounded, and $H_n \downarrow H$ uniformly on compacts.*

These assumptions are fulfilled if for instance the payoff (g_0, g_1) is bounded and H_n is continuous for large enough n. In particular, this holds for pure-cash delivery payoffs $(g_0, 0)$ with bounded and continuous g_0.

Our main result regarding the superhedging price is the following.

Theorem 5.3.6. *Under Assumption 5.3.5, $w_* = w^* = w$ and w is the unique viscosity solution of* (5.19).

The proof will be delegated to the next section. We close this section with a verification result that shows that an optimal replicating strategy can be constructed as long as (5.19) has smooth enough solution, and a remark on how the (5.19) simplifies if the payoff H is a function of the price of risky assets S only.

Remark 5.3.7 (Replicating strategy)**.** Suppose that (5.19) has a smooth enough solution $w \in C^{1,4,2}([0,T] \times \mathbb{R}^d \times \mathbb{R}^d)$. In order to construct Θ satisfying $\Theta_t^{\mathrm{tr}} = w_S(t, \mathcal{S}_t, \mathcal{Y}_t)$ for $\mathcal{S} = \mathcal{S}(S, Y, \Theta)$ and $\mathcal{Y} = \mathcal{Y}(S, Y, \Theta)$, apply Itô's formula to derive an SDE for the triplet $(\Theta, \mathcal{S}, \mathcal{Y})$ with coefficients that are functions of the derivatives of w (up to third order in s) and initial condition $(\Theta_0, \mathcal{S}_0, \mathcal{Y}_0) = (w_S(0, s, y)^{\mathrm{tr}}, s, y)$. If these derivatives are Lipschitz continuous, then existence and uniqueness of $(\Theta, \mathcal{S}, \mathcal{Y})$ would be guaranteed, giving Θ that satisfies (5.17) for $t \in [0, T]$. In particular, using that w is a solution of (5.19) we get for the self-financing portfolio (β, Θ) with $\beta_{0-} = w(0, s, y)$ that $V_T^{\mathrm{liq}}(\Theta) = w(T, \mathcal{S}_T, \mathcal{Y}_T) = H(\mathcal{S}_T, \mathcal{Y}_T)$. Hence, after a block liquidation at time T of the portfolio Θ_T (that does not change V_T^{liq}), leading to new price and impact $\mathcal{S}_T$ and $\mathcal{Y}_T$ respectively, the hedger will be in a position to deliver exactly the physical part of the claim (after a possible new block trade) and have sufficient funds to cover the cash part as well by definition of H.

To summarize, an optimal replicating strategy in this case will have an initial block trade of size $w_S(0, s, y)$ and possibly terminal block trade and will follow a continuous diffusion on $[0, T]$ (to fulfill (5.17)). It requires $w(0, s, y)$ initial capital.

Remark 5.3.8. Let us suppose that H depends only on s and consider the pde

$$-\widetilde{w}_t(t,s) - \tfrac{1}{2} \sum_{i,j=1}^{d} (\Sigma\Sigma^{\mathrm{tr}})_{ij} \partial^2_{s^i s^j} \widetilde{w}(t,s) = 0 \quad \forall (t,s) \in [0,T) \times \mathbb{R}^d, \tag{5.22}$$

with boundary condition $\widetilde{w}(T, s) = H(s)$, where $\widetilde{w} : [0,T] \times \mathbb{R}^d \to \mathbb{R}$. Suppose that (5.22) has a classical solution. Then clearly $\widetilde{w}$ would be a classical solution of (5.19) and hence a viscosity solution as well. Thus, the comparison result in Proposition 5.3.13, see also the proof of Theorem 5.3.6, would imply that $w(t, s, y) = \widetilde{w}(t, s)$ for all $(t, s) \in [0, T] \times \mathbb{R}^d$, or in particular that w does not depend on y. Therefore, the large trader's price of the contingent claim with (cash-equivalent) payoff H equals the small investor's price of the claim (in the multi-asset Bachelier model). Thus, price impact will be irrelevant for the pricing pde, i.e. the price of the contingent claim. Let us note however, that an optimal replicating strategy will be affected by the cross-impact because it will be in the feedback form $\Theta_t^{\mathrm{tr}} = w_S(t, \mathcal{S}_t) = w_S(t, S_t - (\Lambda A + \Gamma)\Theta_t)$.

The function H is independent of y if for instance g_0 and g_1 are functions of s only. Indeed, this follows from the definition of H and because $G^{s - \Lambda y, y^0}(\theta) = \bar{s}^{\mathrm{tr}}\theta + \frac{1}{2}\theta^{\mathrm{tr}}(\Lambda A + \Gamma)\theta$ is independent of y.

The same conclusions will hold true even without assuming that (5.22) has a classical solution by noticing that in this case we do not need $\mathcal{Y}$ as effective coordinate anymore, see also Corollary 4.4.10 and its proof.

5.3.4 Proofs

This section collects all technical results and proofs for the previous sections. First, we show the semi-continuity properties of w_* and w^*.

Lemma 5.3.9. *The functions w_* and w^* are respectively lower and upper semicontinuous. Moreover, $w_*(t,s,y) = \liminf_{(t',s',y')} w(t',s',y')$ for all $(t,s,y) \in [0,T] \times \mathbb{R}^d \times \mathbb{R}^d$, where the limit is taken over $t' < T$. In particular, w_* is the lower semicontinuous envelope of w.*

Proof. We only show that w^* is upper semicontinuous, the lower semicontinuity of w_* follows analogously. Assume by contradiction that $(t_n, s_n, y_n) \to (t,s,y)$ and $w^*(t_n,s_n,y_n) \geq w^*(t,s,y) + \delta$ for some $\delta > 0$ and all n large enough. By definition of w^*, there are $\bar{t}_n, \bar{s}_n, \bar{y}_n$, k_n and n_0 so that $|\bar{t}_n - t_n|, |\bar{s}_n - s_n|, |\bar{y}_n - y_n| < 1/n$, $k_n \to +\infty$ and

$$w_{k_n}(\bar{t}_n, \bar{s}_n, \bar{y}_n) \geq w^*(t_n,s_n,y_n) - \delta/2 \geq w^*(t,s,y) + \delta/2 \quad \text{for } n > n_0.$$

Hence, $\liminf_n w_{k_n}(\bar{t}_n, \bar{s}_n, \bar{y}_n) > w^*(t,s,y)$, a contradiction to the definition of w^* since $(\bar{t}_n, \bar{s}_n, \bar{y}_n) \to (t,s,y)$.

The last claim follows from the fact that the sequence w_k is monotonically decreasing by construction and $w = \downarrow - \lim_{k\to\infty} w_k$. □

In what follows we give the proof of Theorem 5.3.6. We show in the subsequent statements, using the DPP in Theorem 5.3.4, that w_* and w^* are respectively a viscosity supersolution and a subsolution of (5.19). This together with a comparison principle in Proposition 5.3.13 will complete the proof.

Proposition 5.3.10. *The function w_* is a viscosity supersolution of* (5.19).

Proof. The idea of the proof is similar to that of Theorem 4.8.2, but we detail it here for completeness.

Case 1: viscosity property in the interior. First, let $(t_0,s_0,y_0) \in [0,T) \times \mathbb{R}^d \times \mathbb{R}^d$ and $\varphi \in C_b^\infty([0,T] \times \mathbb{R}^d \times \mathbb{R}^d)$ be a smooth function such that

$$\text{(strict)} \quad \min_{[0,T]\times\mathbb{R}^d\times\mathbb{R}^d} (w_* - \varphi) = (w_* - \varphi)(t_0,s_0,y_0) = 0.$$

Suppose that $\mathcal{L}^{[\partial_S \varphi(t_0,s_0,y_0)]}\varphi(t_0,s_0,y_0) < 0$. By continuity of the operator $\mathcal{L}^{[\theta]}$ and the derivatives of φ, there exists a bounded open neighborhood $\mathcal{O} \subset [0,T] \times \mathbb{R}^d \times \mathbb{R}^d$ of (t_0,s_0,y_0) and $\varepsilon > 0$ such that $\mathcal{L}^{[\theta]}\varphi(t,s,y) < -\varepsilon$ in for all $(t,s,y) \in \mathcal{O}$ and $\theta \in \mathbb{R}^d$ with $|\varphi_S(t,s,y) - \theta| \leq \varepsilon$.

Let now $(t_n,s_n,y_n)_n \subset \mathcal{O}$ be such that $(t_n,s_n,y_n) \to (t_0,s_0,y_0)$ with $w(t_n,s_n,y_n) \to$ $\to w_*(t_0,s_0,y_0)$ (here using that w_* is the lower-semicontinuous envelope of w, cf. Lemma 5.3.9), and set $v_n := w(t_n,s_n,y_n) + 1/n$. Since $v_n > w(t_n,s_n,y_n)$, the first part of Theorem 5.3.4 gives the existence of $\theta_n \in \mathbb{R}^d$ and strategies $\gamma_n \in \Gamma$ such that for stopping times τ_n (to be specified later) we have $\mathbb{P}$-a.s.

$$V^{\mathrm{liq},t_n,z_n,\gamma_n}_{t\wedge\tau_n} \geq w(\cdot,\mathcal{S}(S^{t_n,z_n,\gamma_n},Y^{t_n,z_n,\gamma_n},\Theta^{t_n,z_n,\gamma_n}),Y^{t_n,z_n,\gamma_n}-A\Theta^{t_n,z_n,\gamma_n})_{t\wedge\tau_n}, \quad (5.23)$$

where $z_n = (s_n + (\Lambda A+\Gamma)\theta_n, y_n + A\theta_n, \theta_n, v_n)$. To ease the notation in what follows, we will use superscript n instead of superscript (t_n, z_n, γ_n) and

$$\mathcal{S}^n = \mathcal{S}(S^{t_n,z_n,\gamma_n},Y^{t_n,z_n,\gamma_n},\Theta^{t_n,z_n,\gamma_n}), \quad \mathcal{Y}^n = Y^n - A\Theta^n.$$

Take now $\tau_n = \inf\{t \geq t_n \ \ (t,\mathcal{S}^n_t,\mathcal{Y}^n_t)\in\partial_p\mathcal{O}\}$, where $\partial_p\mathcal{O}$ denotes the parabolic boundary of the open set $\mathcal{O}$. In particular, $\tau_n\leq T$. Since $w\geq w_*\geq\varphi$ and $w_*-\varphi$ has a strict local minimum at (t_0,s_0,y_0), there exists $\iota>0$ such that

$$(w-\varphi)(\tau_n,\mathcal{S}^n_{\tau_n},\mathcal{Y}^n_{\tau_n})\geq\iota.$$

Hence, $V^{\mathrm{liq},n}_{\tau_n}-\varphi(\tau_n,\mathcal{S}^n_{\tau_n},\mathcal{Y}^n_{\tau_n})\geq\iota$. Now, (5.16) together with the fact that $\mathcal{S}^n_{t_n}=s_n$, $\mathcal{Y}^n_{t_n}=y_n$ gives that $\mathbb{P}$-a.s.

$$\begin{aligned}\iota &\leq v_n-\varphi(t_n,s_n,y_n)+\int_{t_n}^{\tau_n}\langle\Theta^n_u-\partial_S\varphi(u,\mathcal{S}^n_u,\mathcal{Y}^n_u)^{\mathrm{tr}},\,\mathrm{d}\overline{S}_u\rangle\\ &\quad+\int_{t_n}^{\tau_n}\mathcal{L}^{[\Theta^n_u]}\varphi(u,\mathcal{S}^n_u,\mathcal{Y}^n_u)(\mathbb{1}_{\{|\Theta^n_u-\partial_S\varphi(u,\mathcal{S}^n_u,\mathcal{Y}^n_u)^{\mathrm{tr}}|\leq\varepsilon\}}+\mathbb{1}_{\{|\Theta^n_u-\partial_S\varphi(u,\mathcal{S}^n_u,\mathcal{Y}^n_u)^{\mathrm{tr}}|>\varepsilon\}})\,\mathrm{d}u\\ &\leq v_n-\varphi(t_n,s_n,y_n)+\int_{t_n}^{\tau_n}\langle\Theta^n_u-\partial_S\varphi(u,\mathcal{S}^n_u,\mathcal{Y}^n_u)^{\mathrm{tr}},\,\mathrm{d}\overline{S}_u\rangle\\ &\quad+\int_{t_n}^{\tau_n}\mathcal{L}^{[\Theta^n_u]}\varphi(u,\mathcal{S}^n_u,\mathcal{Y}^n_u)\mathbb{1}_{\{|\Theta^n_u-\partial_S\varphi(u,\mathcal{S}^n_u,\mathcal{Y}^n_u)^{\mathrm{tr}}|>\varepsilon\}}\,\mathrm{d}u. \qquad (5.24)\end{aligned}$$

We would like to perform change of measure that would turn the integral terms in (5.24) into a (stopped) martingale, thus will vanish after taking expectation under the new measure. Our assumption on the structure of $\overline{S}$ gives an equivalent martingale measure for $\overline{S}$ that "kills" the drift terms in the dynamics of $\overline{S}$. So essentially we need to find an equivalent measure under which $\langle a^n_t,\Sigma B_t\rangle+b^n_t\mathbb{1}_{\{|a^n_t|>\varepsilon\}}t$ is a martingale for the bounded $\mathbb{R}^d$-valued process $a^n=\Theta^n-\partial_S\varphi(u,\mathcal{S}^n,\mathcal{Y}^n)$ and real-valued process $b^n=\mathcal{L}^{[\Theta^n]}\varphi(\cdot,\mathcal{S}^n,\mathcal{Y}^n)$. But this is always possible as long as Σ is invertible, being also the case by our assumptions on $\overline{S}$. Hence, we conclude the existence of a measure $\mathbb{P}^n\approx\mathbb{P}$ on $\mathcal{F}_{\tau_n}$ such that

$$\int_{t_n}^{t\wedge\tau_n}\langle\Theta^n_u-\partial_S\varphi(u,\mathcal{S}^n_u,\mathcal{Y}^n_u)^{\mathrm{tr}},\,\mathrm{d}\overline{S}_u\rangle+\int_{t_n}^{t\wedge\tau_n}\mathcal{L}^{[\Theta^n_u]}\varphi(u,\mathcal{S}^n_u,\mathcal{Y}^n_u)\mathbb{1}_{\{|\Theta^n_u-\partial_S\varphi(u,\mathcal{S}^n_u,\mathcal{Y}^n_u)^{\mathrm{tr}}|>\varepsilon\}}\,\mathrm{d}u$$

is a $\mathbb{P}^n$-martingale. Taking expectation under $\mathbb{P}^n$ in (5.24) then gives

$$v_n-\varphi(t_n,s_n,y_n)\geq\iota>0,$$

that holds for all $n\in\mathbb{N}$. However, this is a contradiction since by the choice of v_n and the sequence $(t_n, s_n, y_n)_n$

$$v_n - \varphi(t_n, s_n, y_n) \longrightarrow w_*(t_0, s_0, y_0) - \varphi(t_0, s_0, y_0) = 0.$$

Thus, we have proved the supersolution property of w on $[0,T) \times \mathbb{R}^d \times \mathbb{R}^d$.

Case 2: viscosity property at the boundary.

Let $(s_0, y_0) \in \mathbb{R}_+ \times \mathbb{R}$ and $\varphi \in C_b^\infty([0,T] \times \mathbb{R}^d \times \mathbb{R}^d)$ be a smooth function such that

$$\text{(strict)} \min_{[0,T]\times\mathbb{R}_+\times\mathbb{R}} (w_* - \varphi) = (w_* - \varphi)(T, s_0, y_0) = 0.$$

Suppose that $w_*(T, s_0, y_0) - H(s_0, y_0) < 0$. Then also $\varphi(T, s_0, y_0) - H(s_0, y_0) < 0$. After possibly modifying the test function φ by $(t, s, y) \mapsto \varphi(t, s, y) - \sqrt{T-t}$, we can assume that $\partial_t\varphi(t, s, y) \to +\infty$ when $t \to T$, uniformly on compacts. Hence, in an ε-neighborhood $[T-\varepsilon, T) \times B_\varepsilon(s_0, y_0)$ around (T, s_0, y_0) we have $\mathcal{L}^{[\theta]}\varphi < 0$ for θ in a neighborhood of $\partial_S\varphi(T, s_0, y_0)$. Moreover, after possibly decreasing ε we can assume that $\varphi(T,\cdot) \le H(\cdot) - \iota_1$ on $B_\varepsilon(s_0, y_0)$ for some $\iota_1 > 0$. We argue as in Case 1 above to get (using that $w(T,\cdot) = H(\cdot)$)

$$V^{\mathrm{liq},n}_{\tau_n} - \varphi(\tau_n, \mathcal{S}^n_{\tau_n}, \mathcal{Y}^n_{\tau_n}) \ge \iota_1 \wedge \iota_2,$$

where $\iota_2 := \inf_{[T-\varepsilon,T)\times\partial B_\varepsilon(s_0,y_0)}(w_* - \varphi) > 0$, and a contradiction follows as in Case 1 above. $\square$

Proposition 5.3.11. *The function w^* is a viscosity subsolution of* (5.19).

Proof. Let $\varphi \in C_b^\infty([0,T], \mathbb{R}^d\times\mathbb{R}^d)$ be a test function such that $(t_0, s_0, y_0) \in [0,T]\times\mathbb{R}_+\times\mathbb{R}$ is a strict (local) maximum of $w^* - \varphi$, i.e.

$$\text{(strict)} \max_{[0,T]\times\mathbb{R}_+\times\mathbb{R}} (w^* - \varphi) = (w^* - \varphi)(t_0, s_0, y_0) = 0.$$

Case 1: viscosity property in the interior.

First assume that $t_0 < T$. To ease the notations, we will use the variable x to denote the pair (s, y). Because of the special form of the DPP, Part 2, Theorem 5.3.4, we need to employ w_k (instead of w as we did in the proof for the supersolution property). Lemma 5.3.12 below gives the existence of a sequence $(k_n, t_n, x_n)_{n\ge1}$ such that $k_n \to \infty$, (t_n, x_n) is a local maxima of $w^*_{k_n} - \varphi$, and $(t_n, x_n, w_{k_n}(t_n, x_n)) \to (t_0, x_0, w^*(t_0, x_0))$.

Assume that $\mathcal{L}^{[\partial_S\varphi(t_0,x_0)]}\varphi(t_0, x_0) < 0$ and let $\varphi_n(t,x) = \varphi(t,x) + |t-t_n|^2 + |y-y_n|^2 + {}$ $+ |s-s_n|^4$. Then $\mathcal{L}^{[\partial_S\varphi_n]}\varphi_n > 0$ in a neighborhood B of (t_0, x_0) that contains (t_n, x_n) (for all n large enough). Since our analysis will be restricted to the local neighborhood B, we can modify (in a smooth way) the functions h and φ_n outside of B to be supported on a slightly larger compact set. Thus, (after possibly changing $n \ge 1$) we can construct controls $\gamma_n \in \Gamma_{k_n}$ like in Remark 5.3.7, such that

$$\Theta^{t_n,z_n,\gamma_n}_t = \partial_S\varphi_n(t, \mathcal{S}^{t_n,z_n,\gamma_n}_t, \mathcal{Y}^{t_n,z_n,\gamma_n}_t)^{\mathrm{tr}}, \quad t \ge t_n,$$

where $\mathcal{S}^n_t := \mathcal{S}(S^{t_n,z_n,\gamma_n}_t, Y^{t_n,z_n,\gamma_n}_t, \Theta^{t_n,z_n,\gamma_n}_t)$, $\mathcal{Y}^n_t := Y^{t_n,z_n,\gamma_n}_t - A\Theta^{t_n,z_n,\gamma_n}_t$ and $z_n = (s_n, y_n, 0, w_{k_n}(t_n,x_n) - n^{-1})$.

Let τ_n be the first time after t_n when the process $(t, \mathcal{S}^n_t, \mathcal{Y}^n_t)_{t\geq t_n}$ leaves B. Applying Itô's formula, using (5.16) and $\mathcal{L}^{[\partial_S \varphi_n]}\varphi_n > 0$ on B, we get

$$\begin{aligned} V^{\mathrm{liq},t_n,z_n,\gamma_n}_{\tau_n} &= \varphi_n(\tau_n, \mathcal{S}^n_{\tau_n}, \mathcal{Y}^n_{\tau_n}) + v_n - \varphi_n(t_n,x_n) + \int_{t_n}^{\tau_n} \mathcal{L}^{[\partial_S\varphi_n]}\varphi_n(\mathcal{S}^n_t, \mathcal{Y}^n_t)\,\mathrm{d}t \\ &\geq \varphi_n(\tau_n, \mathcal{S}^n_{\tau_n}, \mathcal{Y}^n_{\tau_n}) + v_n - \varphi_n(t_n,x_n). \end{aligned}$$

Let $2\varepsilon = \inf\{|t-t_0|^2 + |y-y_0|^2 + |s-s_0|^4 \mid (t,s,y) \in \partial B\}$. Then we have

$$\begin{aligned} \varphi_n(\tau_n, \mathcal{S}^n_{\tau_n}, \mathcal{Y}^n_{\tau_n}) &= \varphi(\tau_n, \mathcal{S}^n_{\tau_n}, \mathcal{Y}^n_{\tau_n}) + |\tau_n - t_n|^2 + |\mathcal{Y}^n_{\tau_n} - y_n|^2 + |\mathcal{S}^n_{\tau_n} - s_n|^4 \\ &\geq w_{k_{n-1}}(\tau_n, \mathcal{S}^n_{\tau_n}, \mathcal{Y}^n_{\tau_n}) + |\tau_n - t_n|^2 + |\mathcal{Y}^n_{\tau_n} - y_n|^2 + |\mathcal{S}^n_{\tau_n} - s_n|^4 \\ &\geq w_{k_{n-1}}(\tau_n, \mathcal{S}^n_{\tau_n}, \mathcal{Y}^n_{\tau_n}) + \varepsilon, \end{aligned}$$

where the last inequality holds for all sufficiently large n since $(t_n,s_n,y_n) \to (t_0,s_0,y_0)$. Since also $v_n - \varphi_n(t_n,x_n) \to 0$ as $n\to\infty$, we can find n such that

$$V^{\mathrm{liq},t_n,z_n,\gamma_n}_{\tau_n} > w_{k_{n-1}}(\tau_n, \mathcal{S}^n_{\tau_n}, \mathcal{Y}^n_{\tau_n}). \tag{5.25}$$

Moreover, we can choose the sequence (k_n) in such a way that $k_n \geq 2k_{n-1}+2$. Thus, $v_n = w_{k_n}(t_n,x_n) - 1/n \leq w_{2k_{n-1}+2}(t_n,x_n)$ and hence (5.25) contradicts the second part of Theorem 5.3.4.

Case 2: viscosity property at the boundary: $t_0 = T$.

Let us explain how to adapt the arguments from Case 1 here. Take t_n, x_n, k_n and v_n from Case 1. Consider here the modified test function

$$\varphi_n(t,x) := \varphi(t,x) + \sqrt{T-t} + |y-y_0|^2 + |s-s_0|^4.$$

Since $\partial_t\varphi_n(t,x) \to -\infty$ as $t \to T$, for large enough n we have $\mathcal{L}^{[\partial_S\varphi_n]}\varphi_n \geq 0$ on $[t_n,T) \times B(x_0)$ for some open neighborhood of x_0. Assume by contradiction that $\varphi(T,x_0) > H(x_0) + \eta$ for some $\eta > 0$. Then after possibly restricting to a subset of $B(x_0)$, we have $\varphi_n(T,\cdot) \geq H(\cdot) + \eta$ on $B(x_0)$. Now use the same controls as in Case 1 but for stopping times τ_n being the minimum between T and the first time $(\mathcal{S}^n, \mathcal{Y}^n)$ leaves $B(x_0)$. Then again

$$\varphi_n(\tau_n, \mathcal{S}^n_{\tau_n}, \mathcal{Y}^n_{\tau_n}) \geq \varphi_n(\tau_n, \mathcal{S}^n_{\tau_n}, \mathcal{Y}^n_{\tau_n}) + v_n - \varphi_n(t_n,x_n).$$

This implies that for large n we have

$$\begin{aligned} \varphi_n(\tau_n, \mathcal{S}^n_{\tau_n}, \mathcal{Y}^n_{\tau_n}) \geq{}& w_{k_{n-1}}(\tau_n, \mathcal{S}^n_{\tau_n}, \mathcal{Y}^n_{\tau_n})\mathbb{1}_{\{\tau_n<T\}} + H(\mathcal{S}^n_{\tau_n}, \mathcal{Y}^n_{\tau_n})\mathbb{1}_{\{\tau_n=T\}} \\ &+ \varepsilon\wedge\eta + v_n - \varphi_n(t_n,x_n), \end{aligned}$$

where $2\varepsilon := \inf\{|y-y_0|^2 + |s-s_0|^4 \mid (s,y) \in \partial B(x_0)\}$. Using that $w_n = H_n$ and that

$(H_n)_n$ converges locally uniformly to H (by Assumption 5.3.5), we conclude that for large enough n

$$\varphi_n(\tau_n, \mathcal{S}^n_{\tau_n}, \mathcal{Y}^n_{\tau_n}) > w_{k_{n-1}}(\tau_n, \mathcal{S}^n_{\tau_n}, \mathcal{Y}^n_{\tau_n}).$$

A contradiction now follows as in Case 1. □

For the proof of Proposition 5.3.11 we used the following technical result.

Lemma 5.3.12. *Let $E \subset [0,\infty] \times \mathbb{R}^d \times \mathbb{R}^d$ and $u_k : E \to \mathbb{R}$ be locally uniformly bounded, $B := \overline{B}_\varepsilon(x_0) \cap E$, and assume that $x_0 \in B$ is a strict maximum point for u^* on B, where $u^*(x) := \limsup_{(k,x')\to(\infty,x)} u_k(x')$. Then there exists a sequence $(x_n)_n$ in B and $k_n \to \infty$ with the following property: x_n is a maximum point for $u^*_{k_n}$ on B, where $u^*_{k_n}$ is the upper semicontinuous envelope of w_{k_n}, i.e. $u^*_{k_n} = \limsup_{x'\to x} u_{k_n}(x')$, and*

$$\lim_{n\to\infty} x_n = x_0, \quad \lim_{n\to\infty} u_{k_n}(x_n) \to u^*(x_0).$$

Proof. Since x_0 is the strict maximum point of u^* on B, we have

$$u^*(x) \leq u^*(x_0) \quad \forall x \in B,$$

where the inequality is strict for $x \neq x_0$. Since B is compact and u^*_k is upper semicontinuous, u^*_k has a maximum point x_k in B for which $u^*_k(x) \leq u^*_k(x_k)$ for all $x \in B$. Therefore, $u^*(x_0) \leq \limsup_{k\to\infty} u^*_k(x_k)$. Now take a sequence $k_n \to \infty$ and its corresponding subsequence, that we still denote by (x_k), such that $\limsup_{k\to\infty} u^*_k(x_k) = \lim_{n\to\infty} u_{k_n}(x_k)$. The sequence (x_k) is bounded (since in B) and hence we can extract a subsequence, denoted again by (x_k), that converges to some $\bar{x} \in B$. By definition of $u^*(\bar{x})$ we have

$$u^*(x_0) \leq \limsup_{k\to\infty} u^*_k(x_k) = \lim_{n\to\infty} u_{k_n}(x_k) \leq u^*(\bar{x}).$$

Since x_0 is the strict maximum point of u^*, we deduce equalities everywhere above, in particular $\lim_{n\to\infty} u_{k_n}(x_k) = u(x_0)$ and that $\bar{x} = x_0$. □

We now close this section with a comparison result that will justify the continuity of w and its characterization as the unique viscosity solution of (5.19).

Proposition 5.3.13. *Let $\mathcal{O}$ be an open subset of $\mathbb{R}^m$ and u (resp. v) be a upper-semicontinuous subsolution (resp. lower-semicontinous subsolution) on $[0,T) \times \mathcal{O}$ of*

$$-\partial_t \varphi - \frac{1}{2} trace(\Sigma\Sigma^{tr} D^2\varphi) - \langle B(\cdot, D\varphi), D\varphi\rangle = 0, \tag{5.26}$$

where A is an $m \times m$ matrix, $B : \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}^m$ is Lipschitz continuous and $D\varphi$ (resp. $D^2\varphi$) denote the gradient (resp. Hessian) of φ. Suppose that u and v are bounded and satisfy $u \leq v$ on the parabolic boundary of $[0,T) \times \mathcal{O}$. Then

$$u \leq v \quad \text{on the closure of } [0,T] \times \mathcal{O}.$$

Proof. The proof follows arguments like in the proof of Theorem 4.8.5. We detail them for completeness.

First, we modify the functions u and v by considering $\tilde{u} = e^{\kappa t}u$, $\tilde{v} = e^{\kappa t}v$, where $\kappa > 0$ is fixed. Then $\tilde{u}$ and $\tilde{v}$ are sub-/super-solutions of the pde

$$\kappa\varphi - \partial_t\varphi - \frac{1}{2}\mathrm{trace}(\Sigma\Sigma^{\mathrm{tr}}D^2\varphi) - \langle B(\cdot, e^{-\kappa t}D\varphi), D\varphi\rangle = 0.$$

We prove comparison for the latter that would in particular imply comparison for (5.26). For the ease of notation, we also omit the tildes.

Suppose by contradiction that

$$\sup_{(t,s,y)\in[0,T]\times\mathcal{O}} (u - v)(t, x) > 0.$$

Then we can find $R > 1$ such that

$$\sup_{(t,s,y)\in[0,T]\times\mathcal{O}_R} (u - v)(t, x) > 0,$$

where $\mathcal{O}_R := (-R, R)^m \cap \mathcal{O}$. In particular, there exists $\delta > 0$ and $(t_0, x_0) \in \overline{\mathcal{O}}_R$ such that $(u - v)(t_0, x_0) = \delta > 0$.

Now consider the bounded upper-semicontinuous function

$$\Phi_n(t, x_1, x_2) := u(t, x_1) - v(t, x_2) - \frac{n}{2}|x_1 - x_2|^2,$$

where we use the Euclidean distance $|\cdot|$. By compactness of $[0, T] \times \overline{\mathcal{O}}_R^2$, Φ_n attains its maximum at some $(t^n, x_1^n, x_2^n) \in [0, T] \times \overline{\mathcal{O}}_R^2$ and clearly

$$\Phi_n(t^n, x_1^n, x_2^n) \geq \delta \quad \forall n \in \mathbb{N}. \tag{5.27}$$

By the arguments in the proof of [BLZ16, Lemma 3.11] we have (after possibly passing to a subsequence)

$$n|x_1^n - x_2^n|^2 \to 0 \ \text{ as } n \to \infty. \tag{5.28}$$

An application of Ishii's lemma, as in [CIL92, Theorem 8.3], gives the existence of $(b^n, X^n, Y^n) \in \mathbb{R} \times \mathcal{S}(m) \times \mathcal{S}(m)$, such that with $p^n = n(x_1^n - x_2^n)$

$$(b^n, p^n, X^n) \in \bar{\mathcal{P}}^+_{\mathcal{O}_a} u(t^n, x_1^n),$$
$$(b^n, p^n, Y^n) \in \bar{\mathcal{P}}^-_{\mathcal{O}_a} v(t^n, x_2^n),$$

where X^n and Y^n satisfy

$$\begin{pmatrix} X^n & 0 \\ 0 & -Y^n \end{pmatrix} \leq 3n \begin{pmatrix} I_m & -I_m \\ -I_m & I_m \end{pmatrix}; \tag{5.29}$$

here $\mathcal{S}(m)$ denotes the set of $m \times m$ symmetric non-negative matrices and $I_m \in \mathcal{S}(m)$ is the identity matrix. Using the viscosity property of u and v at (t^n, x_1^n) and (t^n, x_2^n) respectively, we have

$$\begin{aligned}
&\kappa u(t^n, x_1^n) - b_n - \tfrac{1}{2}\mathrm{trace}(\Sigma\Sigma^{\mathrm{tr}}X^n) - \langle B(x_1^n, e^{-\kappa t}p^n), p^n\rangle &&\leq 0,\\
&\kappa v(t^n, x_2^n) - b_n - \tfrac{1}{2}\mathrm{trace}(\Sigma\Sigma^{\mathrm{tr}}Y^n) - \langle B(x_2^n, e^{-\kappa t}p^n), p^n\rangle &&\geq 0.
\end{aligned}$$

Hence

$$\begin{aligned}
0 < \kappa\delta < \kappa(u(t^n, x_1^n) - v(t^n, x_2^n)) \leq&\\
\leq -\tfrac{1}{2}\mathrm{trace}(\Sigma\Sigma^{\mathrm{tr}}Y^n) + \tfrac{1}{2}\mathrm{trace}(\Sigma\Sigma^{\mathrm{tr}}X^n) -&\\
- \langle B(x_2^n, e^{-\kappa t}p^n), p^n\rangle + \langle B(x_1^n, e^{-\kappa t}p^n), p^n\rangle.&
\end{aligned} \tag{5.30}$$

From (5.29), for all $q \in \mathbb{R}^m$

$$q^{\mathrm{tr}}X^n q - q^{\mathrm{tr}}Y^n q \leq 0,$$

which implies that $-\frac{1}{2}\mathrm{trace}(\Sigma\Sigma^{\mathrm{tr}}Y^n) + \frac{1}{2}\mathrm{trace}(\Sigma\Sigma^{\mathrm{tr}}X^n) \leq 0$. Moreover, the Lipschitz property of B with respect to the x argument gives

$$|\langle B(x_1^n, e^{-\kappa t}p^n) - B(x_2^n, e^{-\kappa t}p^n), p^n\rangle| \leq nL \sum_{i,j=1}^{m} |(x_1^{n,i} - x_2^{n,i})(x_1^{n,j} - x_2^{n,j})|,$$

where $x_k^n = (x_k^{n,1}, \ldots, x_k^{n,m})$ for $k = 1, 2$. However, (5.28) in particular implies that this upper bound converges to 0. Thus, we get a contradiction in (5.30) for large n.

□

Proof of Theorem 5.3.6. The viscosity solution property was proven in Propositions 5.3.10 and 5.3.11, while uniqueness and continuity follows from Proposition 5.3.13. Indeed, we showed in the second steps of the proofs of Propositions 5.3.10 and 5.3.11 that $w_*(T, \cdot) \geq H(\cdot)$ and $w^*(T, \cdot) \leq H(\cdot)$, so Proposition 5.3.13 with $\mathcal{O} = \mathbb{R}^d \times \mathbb{R}^d$ gives

$$w^* \leq w_* \quad \text{on } [0, T] \times \mathbb{R}^d \times \mathbb{R}^d.$$

Since by construction $w_* \leq w^*$ on $[0, T] \times \mathbb{R}^d \times \mathbb{R}^d$ we have $w_* = w^*$. Moreover, we also have by construction $w_* \leq w \leq w^*$ on $[0, T) \times \mathbb{R}^d \times \mathbb{R}^d$, so we conclude the equality $w_* = w^* = w$ on $[0, T) \times \mathbb{R}^d \times \mathbb{R}^d$, and hence continuity of w on $[0, T) \times \mathbb{R}^d$. The continuity of w extends to $[0, T] \times \mathbb{R}^d$ since $w(T, \cdot) = H(\cdot)$ by definition of H and $w_*(T, \cdot) = w^*(T, \cdot) = H(\cdot)$ by the conclusions above.

Uniqueness follows by the following standard argument. If u_1 and u_2 are both bounded (discontinuous) viscosity solutions of (5.19), then we just showed that u_1 and u_2 are continuous with $u_1(T, \cdot) = u_2(T, \cdot) = H$. Hence the comparison result gives both $u_1 \leq u_2$ and $u_2 \leq u_1$, and thus the equality $u_1 = u_2$.

□

Bibliography

[AC01] Robert Almgren and Neil Chriss. Optimal execution of portfolio transactions. *J. Risk*, 3:5–40, 2001.

[AFS10] Aurélien Alfonsi, Antje Fruth, and Alexander Schied. Optimal execution strategies in limit order books with general shape functions. *Quant. Finance*, 10(2):143–157, 2010.

[AKS16] Aurélien Alfonsi, Florian Klöck, and Alexander Schied. Multivariate transient price impact and matrix-valued positive definite functions. *Math. Oper. Res.*, 41(3):914–934, 2016.

[Alm12] Robert Almgren. Optimal trading with stochastic liquidity and volatility. *SIAM J. Financial Math.*, 3(1):163–181, 2012.

[AW95] Joseph Abate and Ward Whitt. Numerical inversion of Laplace transforms of probability distributions. *ORSA J. Comput.*, 7(1):36–43, 1995.

[BB04] Peter Bank and Dietmar Baum. Hedging and portfolio optimization in financial markets with a large trader. *Math. Finance*, 14(1):1–18, 2004.

[BBF17a] Dirk Becherer, Todor Bilarev, and Peter Frentrup. Optimal asset liquidation with multiplicative transient price impact. *Appl. Math. Optim.*, 2017.

[BBF17b] Dirk Becherer, Todor Bilarev, and Peter Frentrup. Stability for gains from large investors' strategies in M1/J1 topologies. *Bernoulli*, 2017. to app., arXiv:1701.02167.

[BBF18a] Dirk Becherer, Todor Bilarev, and Peter Frentrup. Approximating diffusion reflections at elastic boundaries. *Electron. Comm. Probab.*, 2018. to app., arXiv:1710.06342.

[BBF18b] Dirk Becherer, Todor Bilarev, and Peter Frentrup. Optimal liquidation under stochastic liquidity. *Finance Stoch.*, 22(1):39–68, 2018.

[BI13] Árpád Baricz and Mourad E. H. Ismail. Turán type inequalities for Tricomi confluent hypergeometric functions. *Constr. Approx.*, 37(2):195–221, 2013.

[Bil99] Patrick Billingsley. *Convergence of probability measures.* Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons, Inc., New York, second edition, 1999.

[BL98] Dimitris Bertsimas and Andrew W. Lo. Optimal control of execution costs. *J. Financial Markets*, 1(1):1–50, 1998.

[BL12] Enzo Busseti and Fabrizio Lillo. Calibration of optimal execution of financial transactions in the presence of transient market impact. *J. Stat. Mech. Theory Exp.*, 2012(09):P09010, 2012.

[BLZ16] Bruno Bouchard, Grégoire Loeper, and Yiyi Zou. Almost-sure hedging with permanent price impact. *Finance Stoch.*, 20(3):741–771, 2016.

[BLZ17] Bruno Bouchard, Grégoire Loeper, and Yiyi Zou. Hedging of covered options with linear market impact and gamma constraint. *SIAM J. Control Optim.*, 55(5):3319–3348, 2017.

[BR17] Tilmann Blümmel and Thorsten Rheinländer. Financial markets with a large trader. *Ann. Appl. Probab.*, 27(6):3735–3786, 2017.

[BSV16] Peter Bank, H Mete Soner, and Moritz Voß. Hedging with temporary price impact. *Math. Financ. Econ.*, pages 1–25, 2016.

[CdL13] Rama Cont and Adrien de Larrard. Price dynamics in a Markovian limit order market. *SIAM J. Financial Math.*, 4(1):1–25, 2013.

[CEK15] Jean-François Chassagneux, Romuald Elie, and Idris Kharroubi. When terminal facelift enforces delta constraints. *Finance Stoch.*, 19(2):329–362, 2015.

[CGMY02] Peter Carr, Hélyette Geman, Dilip B. Madan, and Marc Yor. The fine structure of asset returns: An empirical investigation. *J. Bus.*, 75(2):305–332, 2002.

[CIL92] Michael G. Crandall, Hitoshi Ishii, and Pierre-Louis Lions. User's guide to viscosity solutions of second order partial differential equations. *Bull. Amer. Math. Soc. (N.S.)*, 27(1):1–67, 1992.

[ÇJP04] Umut Çetin, Robert A. Jarrow, and Philip E. Protter. Liquidity risk and arbitrage pricing theory. *Finance Stoch.*, 8(3):311–341, 2004.

[CL95] Louis K. C. Chan and Josef Lakonishok. The behavior of stock prices around institutional trades. *J. Finance*, 50(4):1147–1174, 1995.

[ÇST10] Umut Çetin, H. Mete Soner, and Nizar Touzi. Option hedging for small investors under liquidity costs. *Finance Stoch.*, 14(3):317–341, 2010.

[DI93] Paul Dupuis and Hitoshi Ishii. SDEs with oblique reflection on nonsmooth domains. *Ann. Probab.*, 21(1):554–580, 1993.

[DM82] Claude Dellacherie and Paul-André Meyer. *Probabilities and Potential. B.* North-Holland, Amsterdam, 1982.

[ES05] Felix Esche and Martin Schweizer. Minimal entropy preserves the Lévy property: How and why. *Stochastic Process. Appl.*, 115(2):299–327, 2005.

[FC18] Peter Friz and Ilya Chevyrev. Canonical RDEs and general semimartingales as rough paths. *Ann. Probab.*, 2018. to app.

[FP11] Rüdiger Frey and Ulrike Polte. Nonlinear Black–Scholes equations in finance: Associated control problems and properties of solutions. *SIAM J. Control Optim.*, 49(1):185–204, 2011.

[Fre98] Rüdiger Frey. Perfect option hedging for a large trader. *Finance Stoch.*, 2(2):115–141, 1998.

[FSU17] Antje Fruth, Torsten Schöneborn, and Mikhail Urusov. Optimal trade execution in order books with stochastic liquidity. *Preprint, available at `http://homepage.alice.de/murusov/papers/fsu-optimal_execution_stochastic.pdf`*, 2017.

[GF00] Israel M. Gelfand and Sergei V. Fomin. *Calculus of Variations.* Dover Books on Mathematics. Dover Publications, 2000.

[GH17] Paulwin Graewe and Ulrich Horst. Optimal trade execution with instantaneous price impact and stochastic resilience. *SIAM J. Control Optim.*, 55(6):3707–3725, 2017.

[GHS18] Paulwin Graewe, Ulrich Horst, and Eric Séré. Smooth solutions to portfolio liquidation problems under price-sensitive market impact. *Stoch. Process. Appl.*, 128(3):979–1006, 2018.

[GP16] Nicolae Gârleanu and Lasse Heje Pedersen. Dynamic portfolio choice with frictions. *J. Econom. Theory*, 165:487–516, 2016.

[GP17] Olivier Guéant and Jiang Pu. Option pricing and hedging with execution costs and market impact. *Math. Finance*, 27(3):803–831, 2017.

[GRS11] Selim Gökay, Alexandre F. Roch, and H. Mete Soner. *Liquidity Models in Continuous and Discrete Time*, pages 333–365. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011.

[GZ15] Xin Guo and Mihail Zervos. Optimal execution with multiplicative price impact. *SIAM J. Financial Math.*, 6(1):281–306, 2015.

[HH11] Vicky Henderson and David Hobson. Optimal liquidation of derivative portfolios. *Math. Finance*, 21(3):365–382, 2011.

[HHK92] Ayman Hindy, Chi-Fu Huang, and David Kreps. On intertemporal preferences in continuous time: The case of certainty. *J. Math. Econom.*, 21(5):401–440, 1992.

[HS04] Gur Huberman and Werner Stanzl. Price manipulation and quasi-arbitrage. *Econometrica*, 72(4):1247–1275, 2004.

[JS03] Jean Jacod and Albert N. Shiryaev. *Limit theorems for stochastic processes*, volume 288 of *Grundlehren der Mathematischen Wissenschaften*. Springer-Verlag, Berlin, second edition, 2003.

[Kal02] Olav Kallenberg. *Foundations of modern probability*. Probability and its Applications (New York). Springer-Verlag, New York, second edition, 2002.

[Kar13] Constantinos Kardaras. On the closure in the Emery topology of semimartingale wealth-process sets. *Ann. Appl. Probab.*, 23(4):1355–1376, 2013.

[Kle08] Achim Klenke. *Probability Theory: A Comprehensive Course*. Universitext. Springer London, 2008.

[KMS17] Olga Klein, Ernst Maug, and Christoph Schneider. Trading strategies of corporate insiders. *J. Financial Markets*, 34:48–68, 2017.

[KP96] Thomas G. Kurtz and Philip E. Protter. Weak convergence of stochastic integrals and differential equations. In *Probabilistic models for nonlinear partial differential equations (Montecatini Terme, 1995)*, volume 1627 of *Lecture Notes in Math.*, pages 1–41. Springer, Berlin, 1996.

[KPP95] Thomas G. Kurtz, Étienne Pardoux, and Philip E. Protter. Stratonovich stochastic differential equations driven by general semimartingales. *Ann. Inst. H. Poincaré Probab. Statist.*, 31(2):351–377, 1995.

[KS86] Ioannis Karatzas and Steven E. Shreve. Equivalent models for finite-fuel stochastic control. *Stochastics*, 18(3-4):245–276, 1986.

[KS02] Jan Kallsen and Albert N. Shiryaev. The cumulant process and Esscher's change of measure. *Finance Stoch.*, 6(4):397–428, 2002.

[Kyl85] Albert S. Kyle. Continuous auctions and insider trading. *Econometrica*, 53(6):1315–1335, 1985.

[LA11] Julian Lorenz and Robert Almgren. Mean-variance optimal adaptive execution. *Appl. Math. Finance*, 18(5):395–422, 2011.

[Leb72] Nikolaĭ N. Lebedev. *Special Functions and Their Applications*. Dover Books on Mathematics. Dover Publications, 1972.

[LL07] Pierre-Louis Lions and Jean-Michel Lasry. Large investor trading impacts on volatility. *Ann. Inst. H. Poincaré Anal. Non Linéaire*, 24(2):311–323, 2007.

[LN17] Charles-Albert Lehalle and Eyal Neuman. Incorporating signals into optimal trading. 2017. arXiv:1704.00847v2.

[Løk12] Arne Løkka. Optimal execution in a multiplicative limit order book. *Preprint, London School of Economics*, 2012.

[Løk14] Arne Løkka. Optimal liquidation in a limit order book for a risk-averse investor. *Math. Finance*, 24:696–727, 2014.

[LS13] Christopher Lorenz and Alexander Schied. Drift dependence of optimal trade execution strategies under transient price impact. *Finance Stoch.*, 17(4):743–770, 2013.

[LY05] Hong Liu and Jiongmin Yong. Option pricing with an illiquid underlying asset market. *J. Econ. Dyn. Control*, 29(12):2125–2156, 2005.

[Mar81] Steven I. Marcus. Modeling and approximation of stochastic differential equations driven by semimartingales. *Stochastics*, 4(3):223–245, 1980/81.

[MBEB17] Iacopo Mastromatteo, Michael Benzaquen, Zoltan Eisler, and Jean-Philippe Bouchaud. Trading lightly: Cross-impact and optimal portfolio execution. 2017. arXiv:1702.03838.

[OW13] Anna Obizhaeva and Jiang Wang. Optimal trading strategy and supply/demand dynamics. *J. Financial Markets*, 16:1–32, 2013.

[Pro05] Philip E. Protter. *Stochastic integration and differential equations*, volume 21 of *Stochastic Modelling and Applied Probability.* Springer-Verlag, Berlin, 2005. Second edition. Version 2.1, Corrected third printing.

[PSS11] Silviu Predoiu, Gennady Shaikhet, and Steven Shreve. Optimal execution in a general one-sided limit-order book. *SIAM J. Financial Math.*, 2(1):183–212, 2011.

[PTW07] Guodong Pang, Rishi Talreja, and Ward Whitt. Martingale proofs of many-server heavy-traffic limits for Markovian queues. *Probab. Surv.*, 4:193–267, 2007.

[PW10] Guodong Pang and Ward Whitt. Continuity of a queueing integral representation in the M_1 topology. *Ann. Appl. Probab.*, 20(1):214–237, 2010.

[Roc11] Alexandre F. Roch. Liquidity risk, price impacts and the replication problem. *Finance Stoch.*, 15(3):399–419, 2011.

[RS13] Alexandre Roch and H. Mete Soner. Resilient price impact of trading and the cost of illiquidity. *Int. J. Theor. Appl. Finance*, 16(6):1350037 (27 pages), 2013.

[RW87] L.C.G. Rogers and David Williams. *Diffusions, Markov Processes and Martingales, Vol. II: Itô Calculus.* Cambridge University Press, second edition, 1987.

[RY99] Daniel Revuz and Marc Yor. *Continuous martingales and Brownian motion*, volume 293 of *Grundlehren der Mathematischen Wissenschaften*. Springer, Berlin, 1999.

[Sch16] Torsten Schöneborn. Adaptive basket liquidation. *Finance Stoch.*, 20(2):455–493, 2016.

[Sko56] Anatoli V. Skorokhod. Limit theorems for stochastic processes. *Theory Probab. Appl.*, 1(3):261–290, 1956.

[SL17] Michael Schneider and Fabrizio Lillo. Cross-impact and no-dynamic-arbitrage. 2017. arXiv:1612.07742.

[SS09] Alexander Schied and Torsten Schöneborn. Risk aversion and the dynamics of optimal liquidation strategies in illiquid markets. *Finance Stoch.*, 13(2):181–204, 2009.

[SST10] Alexander Schied, Torsten Schöneborn, and Michael Tehranchi. Optimal basket liquidation for cara investors is deterministic. *Appl. Math. Finance*, 17(6):471–489, 2010.

[ST02] H. Mete Soner and Nizar Touzi. Dynamic programming for stochastic target problems and geometric flows. *J. Eur. Math. Soc. (JEMS)*, 4(3):201–236, 2002.

[SW00] Philipp J. Schönbucher and Paul Wilmott. The feedback effect of hedging in illiquid markets. *SIAM J. Appl. Math.*, 61(1):232–272 (electronic), 2000.

[SZ17] Alexander Schied and Tao Zhang. A state-constrained differential game arising in optimal portfolio liquidation. *Math. Finance*, 27(3):779–802, 2017.

[TWG17] Gerry Tsoukalas, Jiang Wang, and Kay Giesecke. Dynamic portfolio execution. *Management Science*, 2017. to appear.

[Voß17] Moritz Voß. *Dynamic Hedging in Illiquid Financial Markets*. PhD thesis, TU Berlin, 2017.

[Whi02] Ward Whitt. *Stochastic-process limits*. Springer Series in Operations Research. Springer, New York, 2002.

[WZ65] Eugene Wong and Moshe Zakai. On the convergence of ordinary integrals to stochastic integrals. *Ann. Math. Statist.*, 36(5):1560–1564, 1965.

Selbständigkeitserklärung

Ich erkläre, dass ich die vorliegende Arbeit selbständig und nur unter Verwendung der angegebenen Literatur und Hilfsmittel angefertigt habe.

Berlin, den 01. Juni 2018 Todor Bilarev